



Pós-Graduação em Ciência da Computação

***BRUNO NEIVA MORENO***

**REPRESENTAÇÃO E ANÁLISE DE ENCONTROS ESPAÇO-TEMPORAIS  
PUBLICADOS EM REDES SOCIAIS ONLINE**

**Tese de Doutorado**



Universidade Federal de Pernambuco  
posgraduacao@cin.ufpe.br  
[www.cin.ufpe.br/~posgraduacao](http://www.cin.ufpe.br/~posgraduacao)

RECIFE/2016

Bruno Neiva Moreno

**REPRESENTAÇÃO E ANÁLISE DE ENCONTROS  
ESPAÇO-TEMPORAIS PUBLICADOS EM REDES SOCIAIS ONLINE**

Orientadora: Valéria Cesário Times

*Trabalho apresentado ao Programa de Pós-graduação em  
Ciência da Computação do Centro de Informática da Univer-  
sidade Federal de Pernambuco como requisito parcial para  
obtenção do grau de Doutor em Ciência da Computação.*

RECIFE

2016

Catálogo na fonte  
Bibliotecário Jefferson Luiz Alves Nazareno CRB 4-1758

M843r Moreno, Bruno Neiva.  
Representação e análise de encontros espaço-temporais publicados em  
redes sociais online / Bruno Neiva Moreno – 2016.  
158f.: fig., tab.

Orientadora: Valéria Cesário Times.  
Tese (Doutorado) – Universidade Federal de Pernambuco. CIn. Ciência  
da Computação, Recife, 2016.  
Inclui referências e apêndices.

1. Redes sociais online. 2. Redes complexas. 3. Mineração de dados. I.  
Times, Valéria Cesário. (Orientadora). II. Título.

302.3 CDD (22. ed.) UFPE-MEI 2016-174

**Bruno Neiva Moreno**

**Representação e Análise de Encontros Espaço-temporais Publicados  
em Redes Sociais Online**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutora em Ciência da Computação.

Aprovado em: 09/09/2016.

---

**Orientadora: Profa. Dra. Valéria Cesário Times**

**BANCA EXAMINADORA**

---

Profa. Dra. Ana Carolina Brandão Salgado  
Centro de Informática / UFPE

---

Prof. Dr. Vinicius Cardoso Garcia  
Centro de Informática / UFPE

---

Prof. Dr. Luciano de Andrade Barbosa  
Centro de Informática / UFPE

---

Prof. Dr. Fabrício Benevenuto de Souza  
Departamento de Ciência da Computação / UFMG

---

Prof. Dr. Plácido Antonio de Souza Neto  
Instituto Federal do Rio Grande do Norte / Campus Natal

*Dedico esse trabalho ao Autor da vida, ao Alfa e Ômega, o princípio e o fim, o primeiro e o derradeiro, ao verbo que se fez carne: Jesus Cristo.*

## Agradecimentos

A Deus, pelas condições favoráveis e desfavoráveis de toda a vida; pelas ideias colocadas em minha mente; pelas pessoas que conheci durante todo esse período; pelo tempo vivido com minha família no Canadá; pelos resultados obtidos e, finalmente, pela sua Graça, pelo alimento de Sua Palavra e o renovo trazido pela Sua presença, por ter se apresentado de forma tão pessoal mesmo quando eu não confiava. Você é a fonte e o destino, Sua é toda a honra e toda a glória! Obrigado! “*Combati o bom combate, terminei a corrida, guardei a fé.*”<sup>1</sup>.

À Mayara, pelo dom do amor, pela entrega incondicional, por tanta compreensão, por ter assumido responsabilidades que eram minhas, por tantos “nãos” para si transformados em “sim” para mim durante esses 5 anos. Obrigado por sempre ter extraído de mim ânimo e combustível e por sempre ter me mostrado que eu seria capaz. A quem honra, dai honra e você, minha amada, é comigo detentora desse título. Sem você eu não teria conseguido!

À Maria Luiza, o amor da minha vida. Sua existência, no início deste doutorado, serviu-me de demonstração do sustento do Todo Poderoso. Sua presença constante serviu-me de personificação da Graça e mostrou-me que há sentido em pequenas coisas da vida. Você me trouxe alento em tempos de aflição. Suas orações singelas antes de dormir, com poucas, ingênuas e doces palavras acalmaram um coração que, por vezes, estava pesado e cheio da ausência de paz. Jesus lhe usou para preenchê-lo daquela que excede todo o entendimento. Muito obrigado!

À minha mãe, por toda dedicação e que está, hoje, colhendo mais um fruto dessa entrega. Sou o que sou devido a sua garra, a sua força e a sua fé. Obrigado pelas ações e pelas orações! Ao meu pai, pelo orgulho sempre demonstrado, por ter acreditado, pelo estímulo que sempre me deu para que eu terminasse. Aos meus irmãos que eu amo tanto, obrigado pelas palavras de ânimo, pelas conversas e orações. Aos meus sobrinhos que mesmo diante de um tio cheio de promessas, tiveram que se contentar, apenas, com as promessas. Amo vocês demais!

A Osvaldo e Rejany, que me adotaram como filhos. A Renê, que me tem como irmão. O apoio de vocês nesse processo foi fundamental. Obrigado por sempre terem compreendido minha ausência, obrigado por toda ajuda durante todo esse processo, obrigado por tanto discernimento e sabedoria. Essa vitória também é de vocês!

A todos meus familiares, que sempre compreenderam minha ausência, principalmente no último ano do doutorado. Aos meus avós Otávio (*in memoriam*) e Maria Luiza, que sempre acreditaram e investiram em meus estudos e que foram pais que Deus me deu. À minha avó Terezinha, por toda a ajuda, por todas as orações, por toda boa conversa, por todo exemplo. Aos tios e primos, por cada palavra de ânimo, cada boa risada nos nossos almoços dominicais.

Aos amigos em geral, que sempre me ajudaram em momentos de aflição; a Diego e Luiza, que sempre estiveram juntos nessa caminhada, vocês são a representação de Provérbios 17:17 em minha vida; à igreja Sal e Luz, representada pelos pastores e amigos Rômulo e Camila, que

---

<sup>1</sup>Apóstolo Paulo na segunda carta a Timóteo, capítulo 4, versículo 7, Bíblia Sagrada

caminharam comigo durante todo esse tempo, me sustentaram em oração, conselhos, palavras de ânimo e compreensão. Agradeço, especialmente, ao amigo Danillo que nessa reta final foi de fundamental importância para que eu obtivesse a serenidade e equilíbrio necessários.

A todo o pessoal do IFRN-Currais Novos: os alunos, que muitas vezes tiveram que lidar com minha falta de atenção e mau humor; aos colegas, seja os que trocaram aulas para me ajudar, aplicaram provas para mim ou que deram alguma palavra de ânimo na hora do cafezinho; aos diretores, os professores Andreilson e Elionardo, que sempre tentaram me ajudar quando era possível; agradeço, também, aos meus companheiros de viagens semanais que, frequentemente, foram obrigados a escutar minhas inquietações e angústias e tiveram que lidar com minha ansiedade (demonstrada muitas vezes no velocímetro do carro). Fábio, Ítala e Márcia, muito obrigado pelos conselhos, pela compreensão e pelas tantas palavras de força e ânimo!

Não poderia deixar de agradecer aos profissionais e amigos que conheci no Canadá e, também, à família que construí naquele país. Ao Prof. Stan Matwin pelas valiosas contribuições ao meu trabalho [*Professor Matwin, thank you so much for your valuable contributions and mainly for your patience with my English skills*]. Ao Prof. Stênio Fernandes, pela oportunidade a mim concedida quando no projeto CAPES/DFAIT. Aos Misakas por terem prontamente nos recebido e nos prestado o apoio necessário para que pudéssemos nos instalar naquela terra desconhecida. Finalmente, aos irmãos da *Eastview Baptist Church*, especialmente a Risan e Carmélia, Márcio e Alba, Jefferson, Dalva (*in memoriam*) e Teresa. Obrigado pelas caronas, pelo apoio prestado, pela hospitalidade, pelas orações, pelas conversas.

À Valéria minha mais sincera gratidão: por uma convivência tão saudável, pela amizade e pela confiança, por ter me aceitado como aluno de mestrado e, também, de doutorado, por ter aceitado o desafio de trabalhar em uma área diferente do que foi proposto inicialmente, mesmo depois de dois anos de início de doutorado. Obrigado por todos os conselhos: sem eles eu poderia ter desistido ou parado no meio do caminho. Obrigado pelos puxões de orelha: não tenho dúvidas de que eles também me ajudaram a chegar até aqui. Obrigado por ter tido tanta paciência comigo! Agradeço também a Flávio e Marina por tantas vezes, juntamente com você, terem recebido mim e a minha família com muita hospitalidade. Conviver com vocês durante o período do Canadá foi, para mim e minha família, uma grande bênção.

Ao CIn/UFPE como um todo, pelo ambiente de trabalho proporcionado. Aos professores que em muito contribuíram com o meu amadurecimento profissional e acadêmico.

À CAPES e ao CNPq pelo financiamento da bolsa de estudos no Canadá.

Por fim, agradeço, de todo coração, a todos que contribuíram direta ou indiretamente para mais essa vitória.

Meus sinceros agradecimentos!

*É melhor tentar e falhar, que preocupar-se e ver a vida passar. É melhor tentar, ainda que em vão, que sentar-se fazendo nada até o final. Eu prefiro na chuva caminhar, que em dias tristes em casa me esconder. Prefiro ser feliz, embora louco, que em conformidade viver.*

—MARTIN LUTHER KING

## Resumo

O crescente uso de redes sociais online tem feito com que usuários compartilhem, também, informações detalhadas a respeito dos locais que os mesmos frequentam, criando uma ligação entre o mundo físico (o movimento destes usuários no globo) e o mundo virtual (o que eles expressam sobre esses movimentos nas redes). O “*check-in*” é a funcionalidade responsável pelo compartilhamento da localização. Em uma rede social com essa funcionalidade, qualquer usuário pode publicar o local em que o mesmo está em determinado instante de tempo. Esta tese apresenta novas abordagens de análise de redes sociais online considerando as dimensões social, espacial e temporal que são inerentes à publicação de *check-ins* de usuários.

As informações sociais, espaciais e temporais são definidas sob a perspectiva de encontros de usuários, sendo este o objeto de estudo dessa tese. Encontros ocorrem quando duas pessoas (dimensão social), estão em algum local (dimensão espacial), em determinado instante de tempo (dimensão temporal) e decidem publicar esse encontro através de *check-ins*. Além de apresentar um algoritmo para detecção de encontros, é definido um modelo para representação desses encontros. Este modelo é chamado de SiST (do inglês, *SocIal, Spatial and Temporal*) e modela encontros por meio de redes complexas.

Para validar o modelo proposto, foram utilizados dados reais de redes sociais online. Com esses dados, os encontros foram detectados e analisados sob diferentes perspectivas com o objetivo de investigar a existência de alguma lei que governe a publicação dos mesmos, bem como para identificar padrões relativos a sua ocorrência, como padrões temporais, por exemplo. Além disso, as redes construídas a partir do modelo SiST também foram analisadas em termos de suas propriedades estruturais e topológicas. Por meio de redes SiST também foram estudados padrões de movimentação de usuários, como situações em que usuários se movimentam em grupo no globo ou situações em que um usuário é seguido por outros.

**Palavras-chave:** Análise de Redes Sociais Online, Redes Complexas, Mineração de Dados de Redes Sociais

## Abstract

The growing use of online social networks has caused users to share detailed information about the places they visit, resulting on a clear connection between the physical world (i.e. the movement of these users on the globe) and the virtual world (which they express about these movements in the social network). The functionality responsible for sharing location by users is named as “check in”. In a social network with this feature, any user can publish their visited places. This thesis presents new approaches for online social networks analysis considering the social, spatial and temporal dimensions that are implicit in the publication of users check-ins. Social, spatial and temporal information is defined from the perspective of “user encounters”, which is the study object of this thesis. Users encounters occur when two people (social dimension) are somewhere (spatial dimension) in a given time (temporal dimension) and decide to publish this meeting through check-ins. In addition to the algorithm presented for encounters detection, we also defined a model for representation of these encounters. This model is called as SiST (*SocIal, Spatial and Temporal*). The SiST model basically represent encounters by a graph structure.

To validate the proposed approach, we used real data from online social networks. With these data the users encounters were detected and analyzed from different perspectives aiming at investigating the existence of any law governing the publication of encounters and also to identify patterns related to its occurrence, like temporal patterns, for example. Furthermore, the graphs built from SiST model were also analyzed in terms of its structural and topological properties. Through the SiST networks the users movements were studied as well, like in situations in which users move in group or situations where users are followed by other users.

**Keywords:** Analysis of Online Social Networks, Complex Networks, Social Networking Data Mining

## Lista de Figuras

1.1	Funcionalidade de <i>check-in</i> no Facebook. . . . .	21
1.2	Exemplo de uma rede social representada como um grafo. . . . .	22
2.1	Exemplo de locais mostrados em uma aplicação mobile do Facebook. . . . .	30
2.2	Problema das sete pontes de Königsberg: (a) Mapa da cidade de Königsberg à época, na qual cada área da cidade é rotulada com um ponto de cor diferente; (b) o grafo das pontes de Königsberg, representando cada uma das ilhas da cidade como um nó e cada ligação entre esses nós representam as pontes (COMPEAU; PEVZNER; TESLER, 2011). . . . .	34
2.3	Digrafos orientados com três nós. . . . .	35
2.4	Grafo não-orientado com quatro nós e sua matriz de adjacência. . . . .	35
2.5	Distância entre os nós $s$ e $t$ . . . . .	36
2.6	Diâmetro de dois grafos. . . . .	36
2.7	Excentricidade do nó $s$ . . . . .	37
2.8	Raio de $G$ é igual a 2. . . . .	37
2.9	Centralidade de cada nó representada pelo tamanho do nó. . . . .	37
3.1	Categorias de classificação de trabalhos relacionados . . . . .	47
3.2	Palavras mais utilizadas por usuários do Foursquare nas <i>tips</i> (VASCONCELOS et al., 2012) . . . . .	49
3.3	Comportamento rotineiro de usuários em um sábado a noite em Nova Iorque (FERRARI et al., 2011) . . . . .	51
3.4	As 10 categorias mais populares distribuídas em (a) dias de semana e (b) finais de semana (NOULAS et al., 2011) . . . . .	52
3.5	Tipos de grafos construídos a partir de redes do tipo LBSN (BAO et al., 2013). . . . .	52
3.6	Exemplo de grafo tal qual definido em (GEORGIEV; NOULAS; MASCOLO, 2014a). . . . .	53
4.1	Diagrama de atividades do Algoritmo 1. . . . .	59
4.2	Exemplo de <i>check-in</i> publicado originalmente no Foursquare e republicado no Twitter. . . . .	63
4.3	Comparação entre distribuições teóricas e o número de <i>check-ins</i> por locais. . . . .	65
4.4	Comparação entre distribuições teóricas e o número de <i>check-ins</i> por usuários. . . . .	67
4.5	Quantidade de <i>check-ins</i> por mês. . . . .	69
4.6	Quantidade de <i>check-ins</i> por dia da semana. . . . .	70
4.7	Quantidade de <i>check-ins</i> por hora do dia. . . . .	70
4.8	Quantidade de encontros por mês. . . . .	74

4.9	Número médio de <i>check-ins</i> por dia comparado a número de <i>check-ins</i> para os meses de fevereiro, março e abril para a base BKT_CHO. . . . .	75
4.10	Número médio de <i>check-ins</i> por dia comparado a número de <i>check-ins</i> para os meses de fevereiro, março e abril para as bases GOW_CHO e GOW_SCE. . . . .	75
4.11	Locais em torno do centro de convenções de Austin utilizados para realização de encontros. Imagem obtida a partir do Google Maps. . . . .	76
4.12	Percentual de <i>check-ins</i> no Centro de Convenções de Austin. . . . .	77
4.13	Encontros por dia de semana. . . . .	77
4.14	Encontros não sociais por hora do dia. . . . .	78
4.15	Encontros sociais por hora do dia. . . . .	78
4.16	FDA de encontros sociais por hora do dia. . . . .	79
5.1	Visão conceitual das diferentes dimensões de informações presentes em dados de redes sociais. . . . .	83
5.2	Redes construídas considerando-se (a) informações espaciais e (b) informações espaciais e sociais. . . . .	84
5.3	Redes construídas considerando-se informações espaço-temporais para $\Delta t = 30$ min. . . . .	85
5.4	Rede construída considerando-se informações sociais e espaço-temporais para $\Delta t = 15$ min. . . . .	86
5.5	Exemplo de Rede SiST baseada no cenário hipotético da Seção 5.1.1. . . . .	87
5.6	Modelo conceitual representado por um Diagrama de Classes em UML. . . . .	88
5.7	Exemplo de matriz de adjacência espaço-temporal baseado no cenário hipotético da Seção 5.1.1. . . . .	92
5.8	Diagrama de atividades do Algoritmo 2. . . . .	93
5.9	Rede SiST construída a partir da base GOW_SCE considerando $\Delta t = 15$ minutos. . . . .	94
5.10	Diagrama de atividades do Algoritmo 3. . . . .	97
5.11	Diagrama de atividades do Algoritmo 4. . . . .	99
5.12	Diagrama de atividades do Algoritmo 5. . . . .	102
5.13	Distribuição do grau médio das redes SiST não sociais por dia de semana. . . . .	104
5.14	Distribuição do grau médio das redes SiST sociais por dia de semana. . . . .	105
5.15	Diâmetro de redes SiST para $\Delta t = 15$ minutos. . . . .	106
5.16	Distância de redes SiST para $\Delta t = 1$ hora. . . . .	106
5.17	Coefficiente de agrupamento das redes SiST. . . . .	107
5.18	Análise da assortatividade das redes SiST. . . . .	109
5.19	Correlação in-degree e out-degree de redes SiST. . . . .	109
5.20	Histogramas do número de encontros por dia de semana e por hora do dia. . . . .	111
5.21	Mapa apresentando os locais em que o padrão de liderança de encontro descrito foi detectado. . . . .	113

## Lista de Tabelas

2.1	Exemplo de detecção de redes do mundo pequeno (WATTS; STROGATZ, 1998).	41
2.2	Exemplo de Teste de Vuong. . . . .	44
4.1	Bases de Dados utilizadas . . . . .	60
4.2	Quantidade de <i>check-ins</i> de BKT_CHO, GOW_CHO e GOW_SCE . . . . .	61
4.3	Exemplos de <i>check-ins</i> de 4QR_YUA . . . . .	61
4.4	Quantidade de <i>check-ins</i> do 4QR_YUA . . . . .	62
4.5	Quantidade de tweets do TWT_EIN . . . . .	62
4.6	Quantidade de <i>check-ins</i> do TWT_CHE . . . . .	64
4.7	Resumo das informações presentes em cada base de dados. . . . .	64
4.8	Teste de Vuong aplicado sobre dados de <i>check-ins</i> por local. Valores relevantes estão em negrito para auxiliar na interpretação dos dados. . . . .	66
4.9	Teste de Vuong aplicado sobre dados de <i>check-ins</i> por usuários. . . . .	67
4.10	Quantidade de <i>check-ins</i> postados por usuário e local para BKT_CHO, GOW_CHO e GOW_SCE. . . . .	68
4.11	Quantidade de <i>check-ins</i> postados por usuário e local para 4QR_YUA, TWT_EIN e TWT_CHE. . . . .	68
4.12	Teste de Vuong para o número de encontros sociais distribuídos por pares de usuários. . . . .	71
4.13	Teste de Vuong para o número de encontros sociais distribuídos por usuários precursores. . . . .	71
4.14	Teste de Vuong para o número de encontros sociais distribuídos por usuários seguidores. . . . .	72
4.15	Teste de Vuong para o número de encontros sociais distribuídos por locais. . . . .	72
4.16	Relações percentuais de usuários e locais responsáveis pela publicação de mais de 80% encontros. . . . .	73
5.1	Conjunto hipotético de dados de <i>check-ins</i> . . . . .	82
5.2	Conjunto hipotético de dados de amizade . . . . .	83
5.3	Exemplo de <i>check-ins</i> publicados repetidamente. . . . .	95
5.4	Exemplo de <i>check-ins</i> participantes de grupos de encontros. . . . .	97
5.5	Exemplo de <i>check-ins</i> participantes de liderança de encontros. . . . .	100
5.6	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.1. . . . .	107
5.7	Exemplo de padrão de duplicidade de encontros detectado. Check-ins relacionados ao padrão estão em negrito. . . . .	111
5.8	Exemplo de padrão de grupo de encontros detectado. Check-ins relacionados ao padrão estão em negrito. . . . .	112

5.9	Exemplo de padrão de grupo de liderança de encontros detectado. Check-ins relacionados ao padrão estão em negrito. . . . .	112
A.1	Teste de Vuong para o número de encontros não sociais distribuídos por pares de usuários. . . . .	130
A.2	Teste de Vuong para o número de encontros não sociais distribuídos por usuários precursores. . . . .	131
A.3	Teste de Vuong para o número de encontros não sociais distribuídos por usuários seguidores. . . . .	132
A.4	Teste de Vuong para o número de encontros não sociais distribuídos por locais. . . . .	133
B.1	Redes SiST não sociais de BKT_CHO para $\Delta t = 15$ min. . . . .	134
B.2	Redes SiST não sociais de BKT_CHO para $\Delta t = 30$ min. . . . .	134
B.3	Redes SiST não sociais de BKT_CHO para $\Delta t = 1$ hora. . . . .	134
B.4	Redes SiST não sociais de BKT_CHO para $\Delta t = 2$ horas. . . . .	135
B.5	Redes SiST sociais de BKT_CHO para $\Delta t = 15$ min. . . . .	135
B.6	Redes SiST sociais de BKT_CHO para $\Delta t = 30$ min. . . . .	135
B.7	Redes SiST sociais de BKT_CHO para $\Delta t = 1$ hora. . . . .	135
B.8	Redes SiST sociais de BKT_CHO para $\Delta t = 2$ horas. . . . .	136
B.9	Redes SiST não sociais de GOW_CHO para $\Delta t = 15$ min. . . . .	136
B.10	Redes SiST não sociais de GOW_CHO para $\Delta t = 30$ min. . . . .	136
B.11	Redes SiST não sociais de GOW_CHO para $\Delta t = 1$ hora. . . . .	136
B.12	Redes SiST não sociais de GOW_CHO para $\Delta t = 2$ horas. . . . .	137
B.13	Redes SiST sociais de GOW_CHO para $\Delta t = 15$ min. . . . .	137
B.14	Redes SiST sociais de GOW_CHO para $\Delta t = 30$ min. . . . .	137
B.15	Redes SiST sociais de GOW_CHO para $\Delta t = 1$ hora. . . . .	137
B.16	Redes SiST sociais de GOW_CHO para $\Delta t = 2$ horas. . . . .	138
B.17	Redes SiST não sociais de GOW_SCE para $\Delta t = 15$ min. . . . .	138
B.18	Redes SiST não sociais de GOW_SCE para $\Delta t = 30$ min. . . . .	138
B.19	Redes SiST não sociais de GOW_SCE para $\Delta t = 30$ min. . . . .	139
B.20	Redes SiST não sociais de GOW_SCE para $\Delta t = 2$ horas. . . . .	139
B.21	Redes SiST sociais de GOW_SCE para $\Delta t = 15$ min. . . . .	139
B.22	Redes SiST sociais de GOW_SCE para $\Delta t = 30$ min. . . . .	140
B.23	Redes SiST sociais de GOW_SCE para $\Delta t = 1$ hora. . . . .	140
B.24	Redes SiST sociais de GOW_SCE para $\Delta t = 2$ horas. . . . .	140
B.25	Redes SiST não sociais de TWT_CHE para $\Delta t = 15$ min. . . . .	140
B.26	Redes SiST não sociais de TWT_CHE para $\Delta t = 30$ min. . . . .	141
B.27	Redes SiST não sociais de TWT_CHE para $\Delta t = 1$ hora. . . . .	141
B.28	Redes SiST não sociais de TWT_CHE para $\Delta t = 2$ horas. . . . .	141

C.1	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.2.	142
C.2	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.3.	142
C.3	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.4.	142
C.4	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.5.	143
C.5	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.6.	143
C.6	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.7.	143
C.7	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.8.	143
C.8	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.9.	144
C.9	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.10.	144
C.10	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.11.	144
C.11	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.12.	144
C.12	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.13.	145
C.13	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.14.	145
C.14	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.15.	145
C.15	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.16.	145
C.16	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.17.	146
C.17	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.18.	146
C.18	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.19.	146
C.19	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.20.	147
C.20	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.21.	147
C.21	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.22.	147
C.22	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.23.	148
C.23	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.24.	148
C.24	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.25.	148
C.25	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.26.	148
C.26	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.27.	149
C.27	Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.28.	149
D.1	Teste de Vuong para redes SiST não sociais da base BKT_CHO para $\Delta t = 15$ min.	150
D.2	Teste de Vuong para redes SiST não sociais da base BKT_CHO para $\Delta t = 30$ min.	150
D.3	Teste de Vuong para redes SiST não sociais da base BKT_CHO para $\Delta t = 1$ hora.	150
D.4	Teste de Vuong para redes SiST não sociais da base BKT_CHO para $\Delta t = 2$ horas.	151
D.5	Teste de Vuong para redes SiST sociais da base BKT_CHO para $\Delta t = 15$ min.	151
D.6	Teste de Vuong para redes SiST sociais da base BKT_CHO para $\Delta t = 30$ min.	151
D.7	Teste de Vuong para redes SiST sociais da base BKT_CHO para $\Delta t = 1$ hora.	151
D.8	Teste de Vuong para redes SiST sociais da base BKT_CHO para $\Delta t = 2$ horas.	152
D.9	Teste de Vuong para redes SiST não sociais da base GOW_CHO para $\Delta t = 15$ min.	152
D.10	Teste de Vuong para redes SiST não sociais da base GOW_CHO para $\Delta t = 30$ min.	152
D.11	Teste de Vuong para redes SiST não sociais da base GOW_CHO para $\Delta t = 1$ hora.	152

D.12	Teste de Vuong para redes SiST não sociais da base GOW_CHO para $\Delta t = 2$ horas.	153
D.13	Teste de Vuong para redes SiST sociais da base GOW_CHO para $\Delta t = 15$ min. .	153
D.14	Teste de Vuong para redes SiST sociais da base GOW_CHO para $\Delta t = 30$ min. .	153
D.15	Teste de Vuong para redes SiST sociais da base GOW_CHO para $\Delta t = 1$ hora. .	153
D.16	Teste de Vuong para redes SiST sociais da base GOW_CHO para $\Delta t = 2$ horas.	154
D.17	Teste de Vuong para redes SiST não sociais da base GOW_SCE para $\Delta t = 15$ min.	154
D.18	Teste de Vuong para redes SiST não sociais da base GOW_SCE para $\Delta t = 30$ min.	154
D.19	Teste de Vuong para redes SiST não sociais da base GOW_SCE para $\Delta t = 1$ hora.	155
D.20	Teste de Vuong para redes SiST não sociais da base GOW_SCE para $\Delta t = 2$ horas.	155
D.21	Teste de Vuong para redes SiST sociais da base GOW_SCE para $\Delta t = 15$ min. .	155
D.22	Teste de Vuong para redes SiST sociais da base GOW_SCE para $\Delta t = 30$ min. .	156
D.23	Teste de Vuong para redes SiST sociais da base GOW_SCE para $\Delta t = 1$ hora. .	156
D.24	Teste de Vuong para redes SiST sociais da base GOW_SCE para $\Delta t = 2$ horas. .	156
D.25	Teste de Vuong para redes SiST não sociais da base TWT_CHE para $\Delta t = 15$ min.	156
D.26	Teste de Vuong para redes SiST não sociais da base TWT_CHE para $\Delta t = 30$ min.	157
D.27	Teste de Vuong para redes SiST não sociais da base TWT_CHE para $\Delta t = 1$ hora.	157
D.28	Teste de Vuong para redes SiST não sociais da base TWT_CHE para $\Delta t = 2$ horas.	157
D.29	Teste de Vuong para redes SiST não sociais da base 4QR_yua para $\Delta t = 2$ horas.	157
D.30	Teste de Vuong para redes SiST não sociais da base TWT_EIN para $\Delta t = 2$ horas.	158

## Lista de Acrônimos

<b>GPS</b>	Geographical Positioning System
<b>LBSN</b>	Location-based Social Network
<b>LDA</b>	Latent Dirichlet Allocation
<b>OSN</b>	On-line Social Network
<b>SiST</b>	Social, Spatial and Temporal model
<b>SNA</b>	Social Network Analysis
<b>FDA</b>	Função de Distribuição Acumulada
<b>API</b>	Application Programming Interface

## Sumário

<b>1</b>	<b>Introdução</b>	<b>20</b>
1.1	Contextualização . . . . .	21
1.2	Motivação . . . . .	23
1.3	Objetivos . . . . .	25
1.4	Estrutura do documento . . . . .	26
<b>2</b>	<b>Fundamentação Teórica</b>	<b>27</b>
2.1	Redes Sociais Online . . . . .	27
2.1.1	Principais Definições da Área de OSN . . . . .	28
2.1.2	Redes Sociais Online como Objeto de Estudo . . . . .	29
2.1.3	Histórico das Redes Sociais Online . . . . .	31
2.1.4	Evolução das Redes Sociais com Serviço de Localização . . . . .	32
2.2	Grafos e Redes Complexas . . . . .	33
2.2.1	Breve Histórico sobre a Teoria dos Grafos . . . . .	33
2.2.2	Métricas e Propriedades Estruturais de Redes Complexas . . . . .	34
2.2.3	Propriedades Especiais de Redes Complexas . . . . .	39
2.2.3.1	Distribuições de Redes Complexas . . . . .	39
2.2.3.2	Redes Aleatórias: o modelo de Erdős e Rényi . . . . .	40
2.2.3.3	Redes do Mundo Pequeno e os Seis Graus de Separação . . . . .	40
2.2.3.4	Redes de Livre Escala e Conexão Preferencial . . . . .	41
2.3	Inferência de Modelos Estatísticos . . . . .	42
2.3.1	Metodologia de Inferência Estatística . . . . .	42
2.3.2	Exemplo de Seleção de Modelo Estatístico . . . . .	44
2.4	Considerações Finais . . . . .	44
<b>3</b>	<b>Trabalhos Relacionados</b>	<b>46</b>
3.1	Introdução . . . . .	46
3.2	Análise de Fenômenos Espaço-Temporais . . . . .	47
3.3	Análise do Comportamento de Usuários . . . . .	49
3.4	Análise da Mobilidade de Usuários . . . . .	50
3.5	Modelagem de Fenômenos Espaço-Temporais . . . . .	52
3.6	Considerações Finais . . . . .	54
<b>4</b>	<b>Detecção de Encontros</b>	<b>56</b>
4.1	Definições Básicas . . . . .	56
4.2	Algoritmo para Detecção de Encontros . . . . .	57
4.3	Metodologia para Análise das Distribuições Estatísticas . . . . .	58

4.4	Descrição das Bases de Dados . . . . .	60
4.5	Resultados . . . . .	64
4.5.1	Análise da Distribuição dos Check-ins . . . . .	65
4.5.2	Análise Temporal dos Check-ins . . . . .	69
4.5.3	Análise da Distribuição dos Encontros . . . . .	70
4.5.4	Análise Temporal dos Encontros . . . . .	73
4.5.5	Duração dos Encontros . . . . .	77
4.6	Considerações Finais . . . . .	78
<b>5</b>	<b>O Modelo SiST</b>	<b>81</b>
5.1	Introdução . . . . .	82
5.1.1	Um Cenário Hipotético . . . . .	82
5.1.2	Representações de um Encontro Espaço-temporal . . . . .	83
5.2	Modelo Conceitual . . . . .	85
5.3	Modelo Lógico de Dados . . . . .	88
5.4	Modelo Físico de Dados . . . . .	91
5.5	Algoritmos que compõem o Modelo SiST . . . . .	92
5.5.1	Algoritmo de Construção de Redes SiST . . . . .	92
5.5.2	Algoritmo de Detecção de Encontros Duplicados . . . . .	95
5.5.3	Algoritmo de Detecção de Grupo de Encontros . . . . .	96
5.5.4	Algoritmo de Detecção de Liderança de Encontros . . . . .	99
5.6	Validação das Redes SiST . . . . .	102
5.6.1	Análise das Redes SiST . . . . .	103
5.6.1.1	Propriedades das Redes SiST . . . . .	103
5.6.1.2	Investigação sobre leis que governam as conexões das redes SiST . . . . .	107
5.6.2	Detecção de Padrões de Mobilidade de Usuários . . . . .	110
5.7	Considerações Finais . . . . .	114
<b>6</b>	<b>Conclusões</b>	<b>115</b>
6.1	Principais Contribuições . . . . .	115
6.2	Trabalhos Futuros . . . . .	120
	<b>Referências</b>	<b>121</b>
	<b>Apêndice</b>	<b>129</b>
<b>A</b>	<b>Tabelas Adicionais do Teste de Vuong do Capítulo 4</b>	<b>130</b>
<b>B</b>	<b>Propriedades das Redes SiST</b>	<b>134</b>

<b>C</b>	<b>Propriedades das Redes Aleatórias</b>	<b>142</b>
<b>D</b>	<b>Tabelas Adicionais do Teste de Vuong do Capítulo 5</b>	<b>150</b>

# 1

## Introdução

As redes sociais on-line (OSNs, do inglês *On-line Social Network*) são ferramentas que permitem a interação entre pessoas utilizando a web (MISLOVE et al., 2007). As primeiras OSNs permitiam, basicamente, a criação de amizades entre usuários, além da possibilidade de usuários publicarem textos em suas páginas pessoais. Com o desenvolvimento das tecnologias e da própria web, usuários das redes sociais online passaram a poder publicar também outros tipos de informações, além de textos, como imagens, vídeos e, inclusive, a posição geográfica em que o usuário esteve em determinado instante de tempo.

As redes sociais online que foram projetadas para o compartilhamento, exclusivamente, dos locais visitados pelos usuários são chamadas de redes sociais baseadas em localização (LBSNs, do inglês *Location-based Social Networks*) (SCCELLATO et al., 2011). Estas redes ganharam popularidade devido a integração de receptores de GPS (do inglês, *Geographical Positioning System*) aos dispositivos móveis e ao consequente uso do sinal de GPS para publicação dos locais visitados pelos usuários na web. Esta funcionalidade é denominada “*check-in*”.

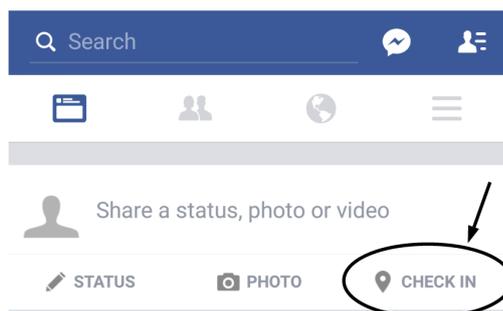
Com a evolução das redes do tipo LBSN, estas passaram a permitir que usuários publicassem também outros tipos de informações integradas ao *check-in*, como algum tipo de mídia (e.g. uma foto do cardápio do restaurante) ou textos (e.g. críticas ou elogios referentes ao restaurante). Do mesmo modo, tornou-se comum que redes de propósito geral, i.e. que não foram originalmente idealizadas como do tipo LBSN, permitissem a publicação de *check-ins*. Este é o caso do Facebook<sup>1</sup>, por exemplo. A Figura 1.1 apresenta a interface *mobile* desta rede social destacando a funcionalidade de *check-in*, incluída desde 2010 (CNN, 2010).

No geral, as redes que permitem compartilhamento de localização combinam as propriedades sociais com as informações geográficas dos usuários (VASCONCELOS et al., 2012). Este serviço tem atraído milhões de pessoas e tem criado, conseqüentemente, uma relação entre o que as pessoas fazem no mundo real e o que publicam no mundo virtual.

As redes sociais online têm criado, portanto, um imenso repositório para análise de informações sociais e espaço-temporais. Esta tese propõe investigar esse repositório de dados sob uma perspectiva diferente do que se tem visto na literatura: a perspectiva do evento denominado *encontro de usuários*, ou simplesmente *encontro*. Um encontro é basicamente definido como

---

<sup>1</sup><http://www.facebook.com>



**Figura 1.1:** Funcionalidade de *check-in* no Facebook.

toda situação em que dois ou mais usuários estão em um mesmo local no mundo real durante o mesmo período e decidem publicar este fato em uma rede social online.

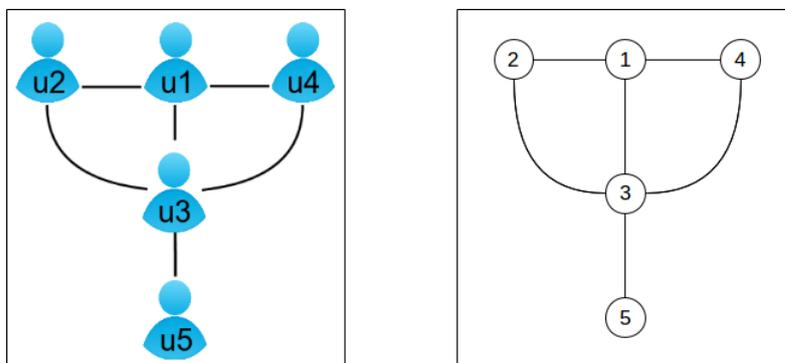
Os encontros publicados pelos usuários de LBSNs são, portanto, o objeto de estudo desta tese. Para estudar esses encontros, as possíveis relações existentes entre eles, as leis que porventura determinam a ocorrência dos mesmos e os padrões inerentes à sua publicação, são propostas abordagens para análise das relações sociais e espaço-temporais dos usuários. Estas abordagens englobam desde a análise completa da probabilidade de dois usuários publicarem um encontro até a definição de um novo modelo para representação de encontros.

Este capítulo está organizado como segue. Primeiramente, a Seção 1.1 contextualiza o leitor no que diz respeito aos conceitos relativos ao objeto de estudo desta tese; a Seção 1.2 descreve as principais motivações desse estudo; na Seção 1.3, os objetivos de pesquisa almejados são apresentados e, por fim, a Seção 1.4 mostra como o referido trabalho está estruturado.

## 1.1 Contextualização

Embora o estudo sobre redes sociais esteja fundamentado em áreas como sociologia (EMIRBAYER; GOODWIN, 1994; MCPHERSON; SMITH-LOVIN; COOK, 2001) e psicologia (KRACKHARDT; PORTER, 1986; GUZZO; SHEA, 1992), recentemente, esses estudos têm focado no fenômeno das redes sociais on-line e seus diferentes tipos (SCHNEIDER et al., 2009; LERMAN; GHOSH, 2010; WANG et al., 2011; YUAN; CONG; SUN, 2014). Além da clara observação das relações sociais demonstradas nas redes, este novo tipo de ambiente permite ainda a construção de um repositório de informações pessoais de forma voluntária pelos seus participantes. Diante disso, a comunidade de pesquisa em geral tem investigado esta grande base de dados por meio da aplicação de diferentes técnicas e realizando análises sobre os dados extraídos de redes sociais on-line, por exemplo: estudos sobre comportamento de usuários (BENEVENUTO et al., 2009; WILSON et al., 2009; ZHUANG et al., 2012; DONG; JIN; KNIJNENBURG, 2015), análise de sentimentos (PANG; LEE, 2008; BERMINGHAM; SMEATON, 2010; TAN et al., 2011; FELDMAN, 2013), análise de influência de usuários em redes sociais (CHA et al., 2010; WU et al., 2011; BAKSHY et al., 2011), recomendação de amigos (BACKSTROM; LESKOVEC, 2011; AGARWAL; BHARADWAJ, 2013), dentre outros.

Membros de uma rede social se relacionam por meio de alguma característica ou atividade em comum do mundo real, podendo ser colegas de trabalho, membros de um mesmo clube ou familiares, por exemplo. Por possuírem algum vínculo, estes membros acabam disseminando e compartilhando informações de diferentes tipos. A Figura 1.2 mostra como uma rede social é comumente representada: por meio de uma estrutura de grafo. Grafos que representam redes sociais são também comumente denominados como um tipo especial de grafo: as redes complexas (WATTS, 2004a). As conexões entre os membros de um grafo como o da Figura 1.2 podem refletir diversos tipos de relacionamentos, e.g. uma relação familiar, de amizade, uma rede profissional, ou outro tipo de relação em que informações são disseminadas e compartilhadas. Normalmente, quando existe conexão entre dois indivíduos de uma rede social online, diz-se que esses dois indivíduos são “amigos” na rede.



**Figura 1.2:** Exemplo de uma rede social representada como um grafo.

O compartilhamento da localização dos usuários em redes sociais online tem impulsionado o uso dessa ferramenta para diferentes propósitos. Redes que possuem essa capacidade têm atraído muita atenção devido ao número de novos serviços que este produto pode disponibilizar, afóra as oportunidades de marketing que a rede oferece aos proprietários de estabelecimentos comerciais. A saber, além de perfis de usuários comuns, ferramentas de LBSNs em geral também permitem que empresas criem seus perfis oficiais. Por exemplo, o proprietário de um restaurante específico pode requerer junto ao Foursquare<sup>2</sup>, a rede do tipo LBSN mais popular da atualidade, que seu restaurante esteja presente na plataforma social por meio de um perfil oficial.

Com o crescimento do uso dessas redes, pesquisas sobre o uso dos dados que elas disponibilizam têm crescido bastante. Pesquisadores têm se interessado em estudar estes dados devido ao forte relacionamento entre o mundo real e o mundo virtual que eles representam. Por exemplo, dados provenientes de redes sociais equipadas com serviços de localização podem ser utilizados para estudos relativos a mobilidade urbana (SCCELLATO et al., 2011; BROWN et al., 2013; GEORGIEV; NOULAS; MASCOLO, 2014b; RIEDERER et al., 2015), para análise do comportamento de usuários em relação aos locais visitados (CRANSHAW et al., 2010; CHENG et al., 2011; SCCELLATO; MASCOLO, 2011), previsão e recomendação de locais a usuários

<sup>2</sup><https://foursquare.com/>

(BACKSTROM; SUN; MARLOW, 2010; YUAN; CONG; SUN, 2014), dentre outros.

Diante das diversas oportunidades de pesquisa relacionadas aos dados sociais e espaço-temporais capturados a partir de redes sociais online, o objetivo desta tese consiste em analisar esses dados sob a perspectiva de encontros de usuários no mundo real que foram publicados nas redes sociais. Para isso, além do estudo sobre a probabilidade de ocorrência de encontros, é proposto aqui um modelo de representação de encontros baseado na teoria dos grafos. Este modelo é definido como modelo SiST (do inglês, *Social, Spatial and Temporal*).

Como o SiST é um modelo de abstração de dados (ELMASRI; NAVATHE, 1999), o mesmo define uma abordagem para representar dados sociais, espaciais e temporais dos usuários e seus respectivos relacionamentos. Em linhas gerais, os grafos gerados a partir do modelo SiST são denominados aqui como redes SiST. As redes SiST se diferenciam dos grafos utilizados em outros trabalhos devido ao fato de que os nós só estão conectados uns aos outros se os usuários que esses nós representam possuem algum tipo de relacionamento espaço-temporal, além do relacionamento social.

Como as redes SiST são baseadas nos conceitos da teoria dos grafos, os métodos e as teorias dessa área de estudo podem ser aplicados para análise das mesmas. Por exemplo, pode-se investigar os encontros modelados por essas redes através de algoritmos de ranqueamento de nós com o objetivo de identificar usuários que possuem um alto poder de influência na publicação de encontros. Pode-se, também, fazer uso de algoritmos de detecção de comunidades para detectar grupos de usuários que estão em determinado local durante um instante de tempo aproximado. Esses grupos de usuários podem ser caracterizados, aqui, como comunidades espaço-temporais.

Além das propriedades e métricas das redes complexas, é possível também investigar o processo de conexão dos nós por meio de uma análise estatística dessas propriedades. Por exemplo, pode-se estudar a distribuição do grau dos nós da rede e descobrir se as conexões entre os nós respeitam uma lei de probabilidade específica, como uma *power-law*, por exemplo.

Para validação das ideias propostas neste documento, foram realizados experimentos com seis bases de dados provenientes de quatro redes sociais distintas. Das seis bases, uma foi capturada do Brightkite e outras duas a partir do Gowalla. Tanto Brightkite como Gowalla foram redes sociais do tipo LBSN bastante populares entre 2009 e 2012, tendo hoje suas atividades encerradas. Além disso, duas bases foram capturadas a partir do Twitter<sup>3</sup> e outra do Foursquare. Por meio desses dados, foi possível identificar qual lei, por exemplo, governa esses encontros e quais padrões temporais caracterizam encontros, além de permitir a descoberta de padrões relativos à mobilidade dos usuários.

## 1.2 Motivação

O pioneiro trabalho de GONZALEZ; HIDALGO; BARABASI (2008) mostrou que a mobilidade de seres humanos possui restrições que vão além das limitações geográficas. De

---

<sup>3</sup><http://www.twitter.com>

acordo com estes autores, as pessoas normalmente frequentam locais que já foram visitados antes por algum conhecido e, inclusive, respeitando certa regularidade temporal. Fazendo-se uso de dados provenientes de redes sociais online, CHO; MYERS; LESKOVEC (2011) reforçaram esta ideia mostrando que além dos fatores espaço-temporais, os relacionamentos sociais explicam cerca de 10% a 30% dos movimentos dos seres humanos.

Ainda de acordo com CHO; MYERS; LESKOVEC (2011), a influência de laços sociais no movimento das pessoas é duas vezes maior do que a influência do próprio movimento para criação de laços sociais. Esses autores mostraram que existe uma probabilidade muito alta de usuários realizarem *check-ins* em algum lugar exatamente depois de amigos terem feito o mesmo. Entretanto, essa probabilidade cai proporcionalmente à medida em que aumenta o intervalo de tempo entre os *check-ins* dos usuários. Do mesmo modo, de acordo com PELECHRINIS; KRISHNAMURTHY (2012), se a influência social é responsável pela similaridade de dois usuários, estes usuários tendem a visitar os locais sugeridos ou visitados por seus amigos.

Mesmo diante das evidências da influência existente entre informações sociais e as informações espaço-temporais, as pesquisas relacionadas ao estudo de eventos que levem em consideração essas três dimensões de informação (i.e. social, espacial e temporal) ainda está em um estágio seminal. Uma breve análise de dados de *check-ins* utilizados nos experimentos dessa tese permitiu a identificação de uma grande quantidade de usuários de redes sociais online que publicam que estão com outros usuários no mesmo local e em um instante de tempo aproximado, i.e. diversos usuários publicam encontros na rede online.

O grande número de usuários que publicam encontros estimulou uma investigação mais detalhada a respeito desses eventos. Através dessa investigação pode-se descobrir que encontros são caracterizados por padrões de publicação diferentes daqueles observados na publicação de *check-ins* e que encontros são governados por leis que se diferenciam, também, das leis que governam a publicação de encontros. Além disso, pode-se notar que é frequente que grupos de usuários publiquem que estão no mesmo local durante o mesmo período de tempo e que determinados usuários tenham um maior poder de influência do que outros usuários no que diz respeito a publicação de encontros. Essas breves descobertas motivaram o trabalho dessa tese no sentido de obter respostas para a seguinte pergunta: “*A publicação de encontros entre usuários de redes sociais online respeita algum padrão de publicação e/ou padrões de mobilidade de usuários?*”. A busca destas respostas está intrinsecamente relacionada com outras perguntas mais específicas, tais como:

P1: De que forma a publicação de *check-ins* se diferencia da publicação de encontros?

P2: Existe alguma lei estatística que modela a publicação de encontros?

P3: Existem fatores temporais que influenciam na publicação de encontros?

P4: Existem padrões de movimentação de usuários que estão, possivelmente, relacionados à publicação de encontros dos mesmos?

P5: Existem pessoas que possuem um maior poder de influência do que outras no que diz respeito à publicação de encontros?

P6: Existem locais que possuem um maior poder de agregação de encontros de usuários?

Para responder a essas perguntas, esta tese considera fatores espaço-temporais e, também, sociológicos, uma vez que para mostrarem que são aceitos em grupos, usuários tendem a compartilhar que estão no mesmo local que outros amigos (DONG; JIN; KNIJNENBURG, 2015). Nesse sentido, o modelo proposto pode ser utilizado para representar e analisar encontros publicados em redes sociais online. Esses encontros foram analisados por meio de dados reais de *check-ins*. Esses dados foram gerados voluntariamente. No total, foram utilizados cerca de 30 milhões de dados de *check-ins*, publicados por aproximadamente 500 mil usuários em cerca de 2 milhões de locais diferentes.

### 1.3 Objetivos

O principal objetivo desta tese é definir abordagens de análise de redes sociais online considerando-se encontros de usuários. Diante disso, relaciona-se as informações sociais, temporais e espaciais modelando-as no formato de redes complexas e estabelecendo algoritmos e métodos que auxiliem na análise desse tipo de dado.

A partir da formalização das interações entre as informações sociais, temporais e espaciais, objetiva-se investigar tanto os padrões de comportamento e uso de redes sociais online com serviços de localização como, também, padrões de mobilidade de usuários. Os objetivos específicos desta tese são sumarizados a seguir.

- Fazer uma descrição da probabilidade de um usuário de uma rede social se encontrar com outro usuário;
- Investigar encontros espaço-temporais de usuários em busca de padrões e leis que governam a publicação desses encontros;
- Definir um modelo de representação de encontros entre usuários;
- Analisar as redes geradas a partir do modelo definido com o objetivo de se detectar padrões de publicação de encontros;

Desta forma, as contribuições desta tese podem ser classificadas como sendo pertencentes a duas grandes áreas: a primeira é técnica, a segunda é sociológica. As contribuições técnicas estão relacionadas à compreensão do motivo da publicação de encontros de usuários de redes sociais online, uma vez que essa compreensão pode auxiliar especialistas a definir novos serviços baseados em localização. As contribuições sociológicas dizem respeito ao fato de que compreender as razões que levam usuários a publicar encontros pode auxiliar na análise de comportamento de usuários em relação à influência de outros usuários e investigar os possíveis impactos dessa influência.

## 1.4 Estrutura do documento

Os capítulos restantes deste documento encontram-se estruturados da seguinte forma:

- **Capítulo 2: Fundamentação Teórica** - Este capítulo descreve os principais conceitos relacionados ao tema dessa tese. Esses conceitos estão distribuídos em duas diferentes áreas: análise de redes sociais e teoria dos grafos;
- **Capítulo 3: Trabalhos Relacionados** - Este capítulo apresenta os trabalhos correlatos à esta tese que foram agrupados em áreas de maior relevância para as contribuições aqui apresentadas;
- **Capítulo 4: Detecção de Encontros** - Este capítulo apresenta uma análise de encontros independentemente da modelagem dos mesmos. Aqui é feita uma descrição da probabilidade de dois usuários se encontrarem em um local e em um intervalo de tempo aproximado. São apresentados padrões de comportamento de usuários relativos a publicação de encontros, padrões temporais e padrões de distribuição de probabilidade desses encontros;
- **Capítulo 5: O Modelo SiST** - Este capítulo descreve o modelo SiST e todos os elementos que o compõem, incluindo os algoritmos para construção de redes SiST e algoritmos para detecção de padrões de mobilidade de usuários. Além disso, este capítulo também apresenta como esse modelo foi validado, seja analisando-se as propriedades das redes construídas, seja por meio da detecção de padrões de mobilidade de usuários.
- **Capítulo 6: Conclusões** - Este capítulo resume as principais contribuições obtidas bem como os trabalhos futuros a serem realizados.

## 2

### **Fundamentação Teórica**

Uma das principais características da Web 2.0 foi transformar o usuário no principal produtor de conteúdo da Internet (O'REILLY, 2005). A partir de meados de 2004, quando o termo “Web 2.0” foi definido, as ferramentas que mais contribuíam com essa transformação foram as redes sociais online. Desde lá, muitas aplicações de redes sociais surgiram e a medida que a tecnologia e os dispositivos móveis foram evoluindo, essas redes acompanharam essa evolução, atraindo até bilhões de usuários, como ocorreu com o Facebook no primeiro semestre de 2016 (STATISTA, 2016a). Todo esse contexto transformou a web em um ambiente repleto de informações publicadas voluntariamente por uma grande quantidade de usuários que, por sua vez, não estão limitados a publicar apenas textos e imagens nas redes sociais online: tem sido cada vez mais comum o compartilhamento também de locais visitados pelos usuários.

Na computação, as redes sociais online são representadas por estruturas denominadas grafos ou, mais especificamente, redes complexas, que são um tipo especial de grafo. Este capítulo tem o objetivo de descrever os principais conceitos da área de redes sociais online, bem como fundamentar as definições relacionadas à teoria dos grafos para o entendimento desta tese. Este capítulo está organizado como se segue, a Seção 2.1 apresenta as redes sociais, descrevendo as principais definições desta área e como se deu a evolução das aplicações de redes sociais online. Na Seção 2.2, os principais conceitos relacionados à teoria dos grafos são listados. A Seção 2.3 descreve a abordagem utilizada ao longo dessa tese para seleção de um modelo estatístico que melhor se ajuste a um conjunto de dados e, finalmente, a Seção 2.4 apresenta as considerações finais deste capítulo.

#### **2.1 Redes Sociais Online**

Esta seção contém as principais definições de redes sociais online (Seção 2.1.1), indicando as principais razões pelas quais esse tipo de ferramenta têm sido amplamente utilizada como objeto de estudo na Computação (Seção 2.1.2) e descrevendo a evolução das redes (Seção 2.1.3). Finalmente, essa seção também apresenta como se deu o desenvolvimento das redes sociais online no que diz respeito ao compartilhamento dos locais visitados pelos usuários (Seção 2.1.4).

### 2.1.1 Principais Definições da Área de OSN

De acordo com MISLOVE (2009), uma rede social online é definida como um sistema onde (a) cada usuário pode criar um perfil dentro da rede, respeitando-se regras de privacidade, ou seja, esse perfil pode ser público ou não; (b) os usuários podem criar conexões com outros usuários que também possuem perfis e, inclusive, também podem se conectar com outros conteúdos; (c) usuários podem navegar dentro da rede. Originalmente, as primeiras redes sociais online permitiam, apenas, que usuários se conectassem entre si e compartilhassem textos e links. Entretanto, essas redes evoluíram ao ponto de permitirem a publicação de diversos tipos de mídias, como imagens e vídeos, por exemplo.

Para contribuir com este cenário, a evolução dos dispositivos móveis, i.e. os *smartphones*, equipados com receptores GPS trouxe uma nova possibilidade em relação ao compartilhamento de informações em redes sociais online: o compartilhamento dos locais visitados pelos usuários. As redes sociais online voltadas para compartilhamento da localização dos usuários foram definidas como redes sociais baseadas em localização, ou LBSN (do inglês, *Location-based Social Network*). A publicação dos locais visitados pelos usuários neste tipo de rede é denominada como uma funcionalidade específica: o “*check-in*”.

As redes sociais online podem ser classificadas pelo tipo de publicação que seus usuários são impulsionados a fazer, apesar de que com a evolução das tecnologias, o propósito de uso dessas redes tem se generalizado bastante (SCCELLATO; MASCOLO, 2011). O Facebook, por exemplo, foi criado com o objetivo de auxiliar usuários a fazer amizades e permitir que os mesmos se conectem uns aos outros, entretanto, hoje o Facebook permite que seus usuários publiquem os mais variados tipos de conteúdo: além de textos e imagens, a rede social online mais popular da atualidade permite também a publicação de vídeos, de transmissão em tempo real, de *check-ins*, além de permitir avaliar pessoas e locais visitados.

No geral, as redes sociais online da atualidade podem ser caracterizadas por possuírem as seguintes funcionalidades e características:

**Perfil do usuário:** o perfil de um usuário pode ser definido, basicamente, como o local onde o usuário se apresenta na rede social online. É no perfil que o usuário descreve informações que possam identifica-lo, como nome, apelido, sexo, o país ou a cidade em que o mesmo reside. As informações do perfil variam entre as aplicações de redes sociais online. Além dessas informações, os perfis dos usuários também podem conter informações sobre gostos e preferências pessoais dos usuários, como visão política, estilo de música preferido ou o status do seu relacionamento afetivo, por exemplo;

**Feed de notícias:** o *feed* é o ambiente pelo qual o usuário acessa o conteúdo exposto na rede. O conteúdo apresentado no *feed* de um usuário não é o mesmo do *feed* de outro usuário, já que o que aparece no *feed* depende dos laços sociais e do perfil de navegabilidade que o usuário possui na rede. O *feed* é também denominado *linha do tempo*;

**Direção da amizade:** as redes sociais podem ser caracterizadas como sendo direcionais ou não direcionais. Uma rede direcional é aquela em que o processo de estabelecimento de amizade é unilateral, i.e. basta que apenas um usuário crie esta relação para que ela seja, de fato, estabelecida. Neste tipo de rede, quando um usuário *a* estabelece um vínculo social com outro usuário *b*, sendo *a* o usuário que criou essa relação, diz-se que *a segue b*, ou seja, *a é seguidor de b*. São exemplos de redes sociais online direcionais: Twitter e Instagram<sup>1</sup>. Neste tipo de rede, o *feed* de um usuário é composto por informações publicadas por aqueles usuários que o mesmo segue. Em redes não direcionais, por sua vez, para que dois usuários estabeleçam laços sociais, os dois usuários precisam concordar com essa conexão. Sendo assim, diz-se que nesses casos os usuários são “amigos” na rede;

**Avaliações:** usuários podem avaliar conteúdo publicado na rede social. Essa funcionalidade se popularizou com o “*like*” do Facebook. As avaliações são comumente utilizadas pelas aplicações de redes sociais como *feedback* do usuário para aprimorar a dinamicidade do *feed*, i.e. baseado no que um usuário avalia na rede, bem como nas informações pessoais apresentadas no perfil, o comportamento do mesmo pode ser avaliado com o objetivo de sugerir publicações, amigos e, inclusive, produtos (CHA; MISLOVE; GUMMADI, 2009);

**Check-in:** é a funcionalidade que permite que em um dado instante de tempo o usuário publique que está visitando determinado local. As redes que apresentam essa funcionalidade normalmente possuem perfis públicos dos locais e suas respectivas coordenadas geográficas. Assim, um usuário que faz uso de uma aplicação de rede social online que possui a funcionalidade de *check-in* é apresentado, normalmente, a um conjunto de locais próximos ao local atual do usuário, conforme ilustrado na Figura 2.1 e, assim, o usuário pode escolher um local e publicar o respectivo *check-in*.

### 2.1.2 Redes Sociais Online como Objeto de Estudo

Embora o estudo sobre as redes sociais não sejam em si uma novidade (RICE, 1927), esses estudos sempre tiveram a limitação relacionada à obtenção dos dados, além do fato de que as redes sociais sempre foram um domínio em que apenas sociólogos, psicólogos e antropólogos se dedicavam a pesquisar (GUZZO; SHEA, 1992; EMIRBAYER; GOODWIN, 1994). Hoje, com as redes sociais online, este cenário tem se modificado bastante.

As mudanças na área de SNA (do inglês, *Social Network Analysis*) têm criado um ambiente multidisciplinar em que as redes sociais online são definidas como o principal objeto de estudo. KURKA; GODOY; ZUBEN (2015) fez um levantamento das principais razões pelas quais as redes sociais online têm atraído tanto o interesse de pesquisadores das mais diversas áreas. Essas razões são sumarizadas a seguir:

---

<sup>1</sup><http://www.instagram.com>



**Figura 2.1:** Exemplo de locais mostrados em uma aplicação mobile do Facebook.

**Disponibilidade de dados:** a quantidade de dados disponibilizados pelas redes sociais online é muito grande. Aliada a essa alta disponibilidade, reside o fato desses dados serem publicados voluntariamente pelos usuários;

**Múltiplos usuários:** redes sociais online permitem produção de conteúdo por diferentes pessoas. Isto faz com que os dados possuam, portanto, diferentes propriedades contextuais, que pode ser desde a linguagem natural do autor da publicação do conteúdo até o estilo de escrita do mesmo, fora o contexto social e político em que o mesmo está inserido;

**Interações:** em redes sociais online, cada indivíduo pode interagir com outro indivíduo e com o conteúdo publicado na rede. Essas interações criam um rico ambiente de estudos;

**Dinâmica temporal:** os dados publicados em redes sociais online são continuamente gerados ao longo do tempo, o que faz com que uma série de transformações sejam perceptíveis, sejam elas estruturais (i.e. que dizem respeito aos relacionamentos entre os usuários) ou relativas ao conteúdo que é publicado;

**Instantaneidade:** além de serem caracterizados por evoluírem com o tempo, os dados de redes sociais são, também, produzidos em qualquer momento e de forma instantânea, ou seja, redes sociais produzem conteúdo em tempo real;

**Ubiquidade:** o que se tem visto é um alto grau de ubiquidade das redes sociais online. Diversos tipos de dispositivos têm sido produzidos para serem conectados às redes sociais online,

sejam eles telefones, televisões ou até óculos e relógios, por exemplo. Essa ideia faz com que a disponibilidade dos dados aumente ainda mais e permita que dados sejam publicados independentemente da localização geográfica do usuário;

Mesmo diante de todas essas razões, a maior dificuldade em se fazer uso dos dados de redes sociais online tem sido o fato de que as principais aplicações de OSN têm se utilizado de protocolos que quando não inibem a captura de seus dados, dificultam bastante: o protocolo OAuth<sup>2</sup> é um exemplo. Diante disso, a rede social online que mais tem sido utilizada nas pesquisas é o Twitter (ROGERS, 2013), uma vez que a política de captura dos dados dessa rede é mais flexível do que em outras redes sociais, como o Facebook, por exemplo.

Em uma pesquisa que se propõe analisar dados de redes sociais online, pode-se elencar diversas consequências provenientes da dificuldade em se obter esses dados. O fato de o usuário de uma rede social ter que permitir que suas informações sejam capturadas por quem quer que seja na web diminui bastante a representatividade dos dados para quem deseja investigar, por exemplo, o comportamento dos usuários em uma rede de uma região específica. Diante disso, muitos pesquisadores têm recorrido à bases de dados já existentes e amplamente utilizadas por outros trabalhos ao invés de construir suas respectivas bases de dados através de *data crawlers* próprios, por exemplo.

### 2.1.3 Histórico das Redes Sociais Online

De acordo com BOYD; ELLISON (2007), o primeiro site reconhecidamente considerado uma rede social online foi o SixDegrees.com, que surgiu em meados de 1997. A partir de então, diferentes tipos de redes sociais online passaram a surgir, essas redes têm basicamente o objetivo de reunir diferentes tipos de pessoas e são voltadas para os mais diversos públicos-alvo. Por exemplo, em 2003 o LinkedIn surgiu com o objetivo de conectar profissionais e organizações, permitindo que usuários mantivessem uma lista de contatos e corporações (LINKEDIN, 2016). O Facebook, por outro lado, surgiu em 2004 como um site de relacionamentos em que o objetivo era, simplesmente, conectar estudantes da Universidade Harvard (BOYD; ELLISON, 2007). Hoje, o LinkedIn possui mais de 430 milhões de usuários (STATISTA, 2016b), enquanto que o Facebook possui mais de 1 bilhão de usuários (STATISTA, 2016a).

Criada em 2006, outra rede social online que se tornou bastante popular foi o Twitter. O Twitter é uma rede social que permite que usuários publiquem mensagens em seus respectivos perfis sendo as mesmas limitadas por 140 caracteres. Uma postagem no Twitter é referenciada como “*tweet*”. O Twitter se tornou popular rapidamente devido a sua característica pública, ou seja, pessoas que não estão cadastradas na rede podem visualizar *tweets* públicos, criando um ambiente de ampla manifestação social a respeito dos mais variados temas (BOYD; GOLDBER; LOTAN, 2010). Em 2012, por exemplo, 100 milhões de usuários postaram 340 milhões de

---

<sup>2</sup><http://oauth.net/>

*tweets* em apenas um dia (TWITTER, 2012). Em março de 2016, o Twitter alcançou a faixa de mais de 310 milhões de usuários cadastrados na ferramenta.

Com a evolução dos dispositivos móveis e a definição de serviços baseados na localização dos usuários, as redes sociais passaram a receber ainda mais atenção com a funcionalidade de *check-in*, permitindo, então, que usuários compartilhassem os locais visitados pelos mesmos. No geral, as redes que permitem compartilhamento de localização combinam as propriedades sociais com as informações geográficas dos usuários (VASCONCELOS et al., 2012). Este serviço tem atraído milhões de pessoas e tem criado, consequentemente, uma relação entre o que as pessoas fazem no mundo real e o que publicam no mundo virtual. A evolução das redes sociais em geral, bem como a popularização dos serviços baseados na localização tem feito com que outros tipos de redes sociais adotem esse tipo de funcionalidade, e não apenas as redes do tipo LBSN.

#### 2.1.4 Evolução das Redes Sociais com Serviço de Localização

De acordo com HUMPHREYS (2007), Dodgeball foi a primeira rede social online a fazer uso do conceito de compartilhamento da localização dos usuários. Esta rede social surgiu em 2003 e pode ser considerada como a primeira rede do tipo LBSN. A rede, entretanto, só permitia que usuários realizassem *check-ins* por meio de SMS. Em 2007 surgiu o Brightkite, uma rede do tipo LBSN que acabou por conquistando ainda mais usuários devido a disponibilização de seus serviços através de diversas plataformas, seja web, SMS e até e-mail (LI; CHEN, 2009). Em 2009 surgiram duas redes do tipo LBSN que passaram a ser bastante popular: o Gowalla e o Foursquare. O Gowalla chegou a ter mais de 600 mil usuários (CAVERLEE et al., 2012), entretanto, com a compra da referida rede pelo Facebook em 2011 e a decisão de descontinuação de seu serviço em 2012, o Foursquare passou a ser a rede social do tipo LBSN mais popular até a atualidade. Um fato que foi preponderante para a finalização do Gowalla foi implementação da funcionalidade de *check-in* no Facebook a partir de 2010 (CNN, 2010).

Não se sabe ao certo quais características mais contribuíram para a popularidade do Foursquare. LINDQVIST et al. (2011) fizeram um levantamento e observaram que o fato da referida rede social oferecer recompensas aos usuários pelos *check-ins* realizados tem sido o maior responsável pela sua popularidade. O aspecto lúdico do Foursquare permite que usuários disputem por pontos, medalhas (*badges*) e prefeituras (*mayorships*) a cada *check-in* realizado. As medalhas são concedidas por diferentes razões, seja por ter feito um *check-in* em um local muito distante, ou seja por ter sido o primeiro a realizar o *check-in* naquele local, por exemplo. Um usuário pode se tornar “prefeito” de um local caso o mesmo seja o usuário que tenha publicado mais *check-ins* nos últimos 60 dias naquele mesmo local. Além do aspecto lúdico, LINDQVIST et al. (2011) observaram que muitos usuários fazem uso da rede social com objetivo de receber tanto as recompensas virtuais como, também, as recompensas do mundo real, afinal, muitas empresas têm utilizado o Foursquare para fornecer descontos e promoções a usuários com base nos *check-ins*, medalhas e prefeituras.

No primeiro semestre de 2014, o Foursquare anunciou a adoção de um novo plano de negócio para a rede social. Esse plano seguiria a tendência de outros aplicativos, dividindo a referida empresa em duas: o próprio Foursquare e uma nova marca, o Swarm (FERREIRA FILHO; ACHER; BARAIS, 2015). Neste caso, Foursquare seria voltado apenas para aqueles usuários que têm o interesse de explorar o local em que o mesmo está em determinado instante e o Swarm voltado para publicação de *check-ins* (WILKEN, 2016).

## 2.2 Grafos e Redes Complexas

Grafos têm sido amplamente utilizados para representação de diferentes conceitos no mundo real, seja na sociologia, na engenharia ou na ciência da computação. As redes complexas, por sua vez, trouxeram ainda mais atenção para a teoria dos grafos após surgimento da “nova ciência das redes” (WATTS, 2004a). Em linhas gerais, uma rede complexa pode ser definida como um tipo especial de grafo, que possui uma grande quantidade de nós e propriedades específicas. Esta seção tem o objetivo apresentar os principais conceitos relativos a grafos e redes complexas e que fundamentam o modelo proposto nessa tese. Esta Seção está organizada como se segue. Na Seção 2.2.1 é apresentado um breve histórico sobre esta área de estudo. Na Seção 2.2.2 são listadas as principais definições da teoria dos grafos bem como as métricas e propriedades que podem ser extraídas a partir desse tipo de estrutura. A Seção 2.2.3 descreve algumas propriedades especiais, tipos diferentes de grafos, bem como fenômenos que podem ser detectados a partir de grafos e redes complexas.

### 2.2.1 Breve Histórico sobre a Teoria dos Grafos

Os grafos são estruturas bastante utilizadas na matemática e na computação. O primeiro resultado da teoria de grafos que se tem conhecimento é de 1736, quando Leonhard Euler formulou o problema das sete pontes de Königsberg (EULER, 1736). O problema das pontes era fruto de uma lenda popular, na qual discutia-se a possibilidade de atravessar todas as pontes de uma determinada região composta por diversas ilhas contanto que cada ponte fosse atravessada somente uma vez e que o ponto de chegada fosse exatamente o ponto de partida. A Figura 2.2 ilustra o problema formalizado pelo primeiro grafo da história em que Euler demonstrou a impossibilidade de tal feito. Neste grafo, cada nó representa uma ilha e as ligações entre os nós representam as pontes. Para que se conseguisse atravessar todas as pontes conforme especificado, Euler sugeriu que cada nó do grafo tivesse um par de ligações. Ou seja, cada ilha deveria conter duas pontes, uma para entrar na ilha e outra para sair.

Muitos anos se passaram e a teoria dos grafos passou a ser utilizada por diversas ciências. Entretanto, a falta de poder de processamento e de suporte computacional era um grande problema para modelagem e análise de grafos complexos, com grandes quantidade de nós e ligações. Apenas neste novo milênio, impulsionados pelos avanços tecnológicos, cientistas das mais diversas áreas realizaram importantes progressos em problemas que até então não eram



**Figura 2.2:** Problema das sete pontes de Königsberg: (a) Mapa da cidade de Königsberg à época, na qual cada área da cidade é rotulada com um ponto de cor diferente; (b) o grafo das pontes de Königsberg, representando cada uma das ilhas da cidade como um nó e cada ligação entre esses nós representam as pontes (COMPEAU; PEVZNER; TESLER, 2011).

solucionados (ALBERT; JEONG; BARABASI, 2000; GONZALEZ; HIDALGO; BARABASI, 2008), além de terem reformulado antigas ideias e introduzido novas técnicas na área de análise de redes (WATTS, 2004b). Este importante momento é exatamente o que WATTS (2004a,b) descreve como “nova ciência das redes”, referindo-se a um grafo como sendo uma “rede complexa” que, dentre outras características, possui milhares de nós.

Apesar de ser um subproduto da teoria dos grafos, as redes complexas só vieram a se popularizar a partir dos estudos da nova ciência das redes. Esse tipo de representação vem sendo utilizado frequentemente para situações em que indivíduos ou objetos se relacionam. Os conceitos básicos relativos às redes complexas, bem como as principais propriedades desse tipo de estrutura são apresentados na Seção 2.2.2.

### 2.2.2 Métricas e Propriedades Estruturais de Redes Complexas

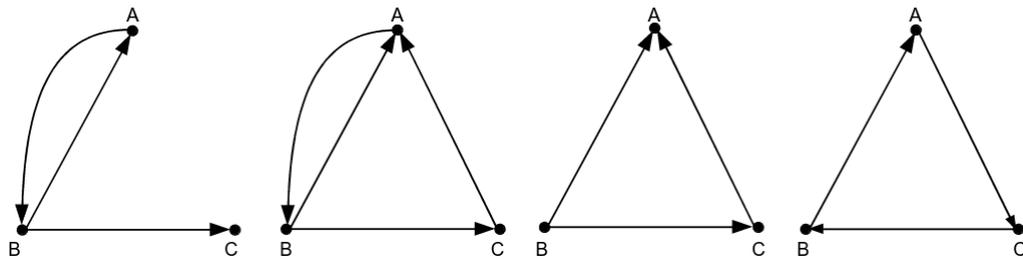
A teoria dos grafos define propriedades e métricas de grafos e algoritmos que podem ser executados sobre esses grafos. As propriedades definidas na teoria dos grafos são frequentemente utilizadas para analisar o grafo como um todo, seja por meio da análise estatística dessas propriedades, da comparação entre diferentes grafos ou da execução de algoritmos que analisam o grafo sob variados pontos de vistas (CHAKRABARTI et al., 1998; PAGE et al., 1998; KLEINBERG, 1999; KLEINBERG; LAWRENCE, 2001).

Uma rede complexa é normalmente definida como um tipo especial de grafo que possui uma grande quantidade de nós WATTS (2004a,b). Basicamente, esse termo vem sendo utilizado para denotar um grafo desde que a nova ciência das redes começou a fazê-lo. Esta seção, assim como esta tese como um todo, faz uso de ambos os termos para representar o mesmo tipo de estrutura: aquela em que nós são conectados uns com os outros através de conexões, ou arestas. Esta seção apresenta conceitos amplamente definidos pela teoria dos grafos.

A conectividade de um grafo é um dos conceitos básicos da teoria dos grafos. Diz-se que um grafo  $G$  é *conectado* quando existe um caminho de qualquer nó do grafo para qualquer outro nó do mesmo grafo. Por outro lado,  $G$  é dito *desconectado*, ou *não conectado*, se existem dois

ou mais nós no grafo sem que qualquer caminho permita alcançá-los (HARARY, 1969).

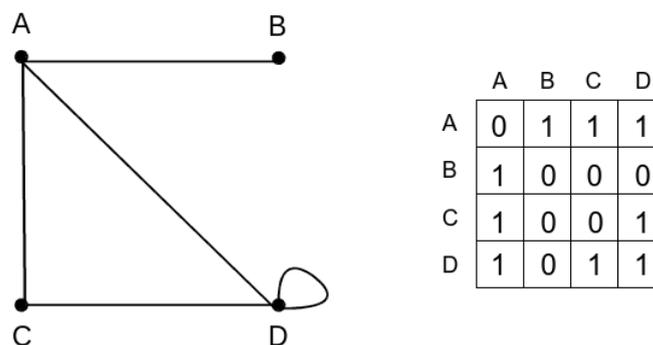
Grafos também podem ser classificados quanto à orientação das arestas entre os nós. Um grafo  $G(V,A)$  é *orientado*, por exemplo, quando as arestas  $A$  são direcionadas. Esse tipo de grafo também é chamado de grafo *direcionado*, *grafo dirigido* ou ainda *digrafo*. Um grafo é dito *bidirecional* quando as conexões são direcionadas para os dois nós pelos quais essas conexões estão sendo conectadas. A Figura 2.3 mostra quatro grafos direcionados, cada qual com três nós.



**Figura 2.3:** Digrafos orientados com três nós.

Um grafo direcionado difere de um grafo não-direcionado pelo fato de que este último não possui as arestas direcionadas. Em um grafo não-direcionado  $G$ , dois nós  $u$  e  $v$  estão conectados sem que haja direção na conexão destes nós.

Além da representação gráfica, grafos também podem ser representados por meio de matrizes. A matriz de um grafo é chamada de matriz de adjacência, ou também matriz de conexão. Uma matriz de adjacência é uma matriz quadrada na qual cada linha e coluna representa um nó do grafo. Cada entrada da matriz indica se os nós relacionados pelas linhas e colunas estão conectados ou não. Ou seja, o valor de uma célula  $a_{ij}$  de uma matriz de adjacência  $M$  pode ser 1, caso os nós representados por  $i$  e  $j$  estejam conectados ou 0, caso contrário. A Figura 2.4 mostra um grafo com quatro nós não-direcionados e a respectiva matriz de adjacência seguindo a referida abordagem.



**Figura 2.4:** Grafo não-orientado com quatro nós e sua matriz de adjacência.

Finalmente, a teoria dos grafos define diversas propriedades de redes complexas. A seguir são descritas as propriedades, métricas e conceitos principais que se relacionam com esta tese. Estudos mais aprofundados sobre o tema foram disponibilizados por MILGRAM (1967), HARARY (1969), BUCKLEY; HARARY (1990), SKIENA (1990), DIESTEL (1997), BARABASI; ALBERT (1999) e WEST (2001).

**Distância entre dois nós:** a distância  $d(s, t)$  refere-se ao custo de travessia entre os nós  $s$  e  $t$  de um grafo. Esse custo é calculado em termos da soma dos pesos de cada aresta percorrida (BUCKLEY; HARARY, 1990). Por vezes essa propriedade é também chamada de “menor caminho” entre dois nós. A Figura 2.5 ilustra dois grafos em que pode-se observar diferentes valores de distância entre os nós  $s$  e  $t$ . Um dos menores caminhos encontrados para cada um dos grafos estão com suas arestas destacadas;

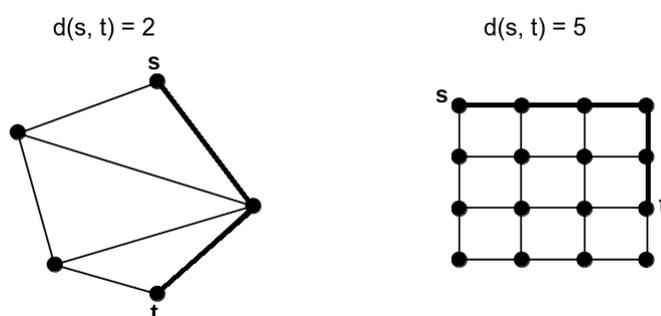


Figura 2.5: Distância entre os nós  $s$  e  $t$ .

**Diâmetro de um grafo:** o diâmetro  $D$  de um grafo  $G$  consiste na maior distância entre dois nós quaisquer de  $G$  (DIESTEL, 1997). Em outras palavras, o diâmetro representa o maior caminho de menor distância entre dois nós. Um grafo desconectado possui um diâmetro infinito (WEST, 2001). A Figura 2.6 mostra dois grafos cujos valores de diâmetros são iguais a 5 e 7, respectivamente. As arestas do caminho que representa a distância mais longa de cada um dos grafos estão em destaque nesta figura;

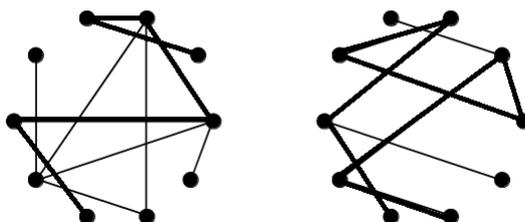


Figura 2.6: Diâmetro de dois grafos.

**Excentricidade de um nó:** a excentricidade de um nó  $s$  de um grafo  $G$  representa a *distância máxima* entre  $s$  e todos os nós de  $G$  (SKIENA, 1990). Para um grafo desconectado, todos os nós são definidos como tendo uma excentricidade infinita. O valor máximo que a excentricidade de um nó pode ter é igual ao valor do diâmetro do grafo como um todo. A Figura 2.7 exhibe um grafo em que as arestas em destaque formam um dos caminhos da excentricidade de  $s$ , que é, portanto, igual a 3;

**Raio de um grafo:** o raio  $r$  de um grafo  $G$  representa a *excentricidade mínima* de qualquer vértice de  $G$  (HARARY, 1969). Um grafo desconectado tem, portanto, um raio infinito

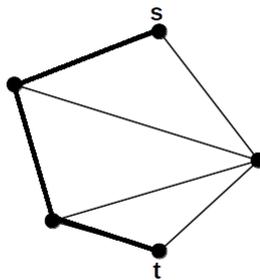


Figura 2.7: Excentricidade do nó  $s$ .

(WEST, 2001). A Figura 2.8 mostra um exemplo de um grafo em que a menor excentricidade encontrada é igual a 2, ou seja, o raio é igual a 2. Há dois caminhos que representam essa distância, são eles  $s \rightarrow t$  e  $s \rightarrow u$ ;

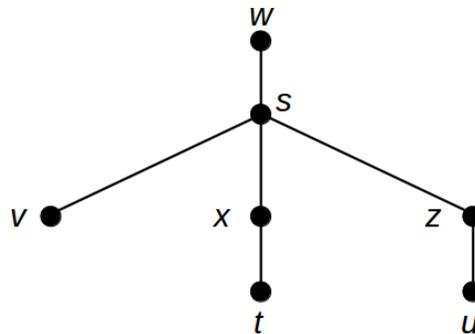


Figura 2.8: Raio de  $G$  é igual a 2.

**Média do menor caminho:** a média  $\sigma d$  do menor caminho de um grafo  $G$  diz respeito a média das distâncias entre os pares de nós de  $G$ . O valor  $\sigma d$  para o grafo à esquerda da Figura 2.5 é igual a 1.75;

**Centralidade de grau de um nó:** a centralidade de grau  $b$  de um nó  $s$  do grafo  $G$  representa quão próximo  $s$  está de qualquer outro nó de  $G$  levando em consideração as conexões incidentes sobre  $s$ . A Figura 2.9 mostra um exemplo em que, quanto maior o tamanho do nó, maior é o valor da centralidade desse nó. Portanto, pode-se observar que o nó representado pelo maior círculo, de fato, possui uma maior incidência de conexões.

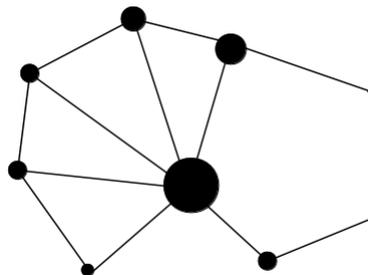


Figura 2.9: Centralidade de cada nó representada pelo tamanho do nó.

Em redes sociais, a centralidade é um dos conceitos mais estudados (BORGATTI, 2005).

Essa propriedade pode ser calculada de acordo com diferentes medidas, dentre elas destacam-se três: a centralidade de proximidade (*closeness*) (FREEMAN, 1978), a centralidade de intermediação (*betweenness*) (FREEMAN, 1978) e a centralidade de autovetor (*eigenvector*) (BONACICH; LLOYD, 2001). Basicamente, o valor da centralidade de proximidade de um nó é dado pela menor distância total desse nó em relação a todos os outros nós da rede. A centralidade de intermediação, por sua vez, atribui importância a um nó em função da passagem de fluxo por ele para interagir com outros nós da rede. Por fim, a centralidade de autovetor atribui relevância a um nó por ele estar conectado a outros nós que se encontram em uma posição mais central da rede.

**Coefficiente de agrupamento:** o coeficiente de agrupamento de um grafo, ou *coeficiente de agrupamento global*, ou ainda *transitividade*, mede a tendência com que os nós de um grafo tendem a agrupar-se WATTS; STROGATZ (1998), i.e. indica se o grafo tende a criar grupos coesos de nós ou não. O coeficiente de agrupamento global  $C_G$  do grafo  $G$  é uma medida dos números de grafos triângulos encontrado em  $G$  e é calculado com base no *coeficiente de agrupamento local* para cada nó  $i$  de  $G$ . O coeficiente de agrupamento local de um nó  $i$  mede o quão próximo os vizinhos de  $i$  estão de serem conectados. O coeficiente de agrupamento local mede a densidade das conexões da vizinhança de  $i$ .

**Assortatividade:** o grau de assortatividade mede a tendência de um nó se conectar com outros nós similares a ele (ABRAMSON et al., 2014). Esta medida pode ser calculada utilizando uma medida de correlação, como o coeficiente de Pearson, por exemplo (YOOK; RADICCHI; MEYER-ORTMANN, 2005). O coeficiente de assortatividade  $r$  de um nó  $i$  sempre tem um valor entre  $-1$  e  $1$ . Caso  $r$  seja igual a  $0$ ,  $i$  cria conexão com outros nós de forma aleatória. Quanto mais próximo  $r$  for de  $1$  ( $r > 0$ ), então  $i$  tende a criar conexões apenas com nós similares (i.e. com grau semelhante). Por outro lado, se  $r < 0$ ,  $i$  tende a criar conexões com nós diferentes, i.e. com nós de diferente grau.

NEWMAN (2003a) apresentou uma série de valores de assortatividade calculados para diferentes grafos construídos a partir de dados reais. Na área de redes sociais, por exemplo, o autor demonstrou que a assortatividade pode variar entre  $-0.029$  e  $0.363$ , dependendo do tipo de relação social representada.

Finalmente, como grafos podem não ser completamente conectados, pode-se analisar esse tipo de estrutura em termos da componente principal do grafo. De acordo com (CORMEN et al., 2001), a componente principal do grafo  $G$ , ou a componente fortemente conectada, é o conjunto máximo de vértices  $C$  de  $G$  tal que para todo par de vértice  $u$  e  $v$  em  $C$ , existe uma conexão mútua entre ambos. A componente principal de um grafo orientado pode ser analisada, também, em termos de componentes principais de entrada e de saída.

### 2.2.3 Propriedades Especiais de Redes Complexas

A Seção 2.2.3.1 lista conceitos relacionados à distribuição de redes complexas. A Seção 2.2.3.2 descreve, brevemente, o modelo de Erdős e Rényi, que é capaz de criar redes aleatórias. Na Seção 2.2.3.3, são discutidos fenômenos especiais relacionados às redes complexas: as redes do mundo pequeno e o fenômeno conhecido como seis graus de separação. Finalmente, a Seção 2.2.3.4 apresenta as redes de livre escala e o fenômeno que demonstra que determinados nós tendem a se conectar com outros nós de acordo com critérios estabelecidos: o fenômeno de conexão preferencial.

#### 2.2.3.1 Distribuições de Redes Complexas

Um tipo comum de análise de redes complexas é a análise da distribuição do grau da rede. A análise de redes do mundo real permitiu que diversos autores concluíssem que a distribuição do grau dessas redes segue, normalmente, uma distribuição *power-law* (BARABASI; ALBERT, 1999; FALOUTSOS; FALOUTSOS; FALOUTSOS, 1999; KUMAR et al., 1999; ADAMIC; BUYUKKOKTEN; ADAR, 2003; BOCCALETTI et al., 2006). Em uma rede governada por esse tipo de distribuição, a maioria dos nós possuem um grau pequeno, i.e. são conectados a poucos nós, enquanto que alguns poucos nós possuem alto grau de conectividade. O fato de uma *power-law* ser uma distribuição do tipo *heavy-tailed* (i.e. de cauda pesada), faz com que existam poucos nós considerados *hubs*, porém existam com uma grande quantidade de conexões quando comparados ao restante da rede. *Hubs* são nós com uma grande quantidade de conexão e que possuem um importante papel no fluxo de informações de uma rede complexa (KLEINBERG, 1999; KLEINBERG; LAWRENCE, 2001). Esses nós possuem muito poder na rede pois são nós que, quando retirados da mesma, podem torná-la desconectada.

Redes governadas por uma *power-law* apresentam algumas propriedades que foram elencadas por MITCHELL (2009) e são descritas a seguir:

- Possuem um baixo número de nós atuando como *hubs* quando comparadas com redes que são distribuídas de acordo com outro tipo de lei que não seja uma *power-law*;
- Seus nós são bastante heterogêneos quanto ao valor dos respectivos graus;
- São redes bastante resilientes, i.e. mesmo removendo nós aleatoriamente, suas propriedades tendem a ser preservadas, uma vez que a maioria dos nós não são *hubs*.

O fato de redes sociais serem, comumente, distribuídas de acordo com uma *power-law* condiz com o que se vê no mundo real: enquanto poucas pessoas têm uma grande quantidade de relacionamentos (i.e. possuem um alto grau), a maioria das pessoas têm poucas relações sociais. Do mesmo modo, a chance de uma pessoa conhecer uma pessoa “popular” (i.e. um *hub*) é maior do que conhecer uma pessoa com poucas conexões.

### 2.2.3.2 Redes Aleatórias: o modelo de Erdős e Rényi

Uma rede aleatória é um grafo no qual os números de nós e de arestas são determinados aleatoriamente. Estudos sobre esse tipo de rede foram iniciados na década de 50, quando Paul Erdős e Alfred Rényi (ERDÖS; RÉNYI, 1960) definiram a primeira abordagem de criação de grafos aleatórios. No decorrer dos anos, e mais especialmente com o surgimento da nova ciência das redes, cientistas descobriram que muitos dos fenômenos naturais poderiam ser descritos em termos de redes aleatórias (BARABASI; ALBERT, 1999; WATTS; STROGATZ, 1998).

O modelo de Erdős e Rényi assume que a construção de uma rede deve-se iniciar definindo-se o número de nós que a rede deve conter e, assim, esses nós serão conectados seguindo uma probabilidade  $p$  que, por sua vez, deve seguir uma distribuição de Poisson ou uma distribuição normal (BARABASI; ALBERT, 1999). Nós de uma rede aleatória muito grande (i.e. com uma grande quantidade de nós) possuem, aproximadamente, o mesmo número de links: grafos criados randomicamente possuem características semelhantes uns com os outros. As ligações entre os nós de uma rede aleatória são criadas de forma independente uma das outras.

O que se tem visto hoje na literatura é que, na verdade, redes sociais do mundo real (BARABASI; ALBERT, 1999) não são aleatórias. Nós dessas redes (i.e. os usuários) não tendem a ter o mesmo número de conexões, uma consequência do fato de que as conexões em redes sociais tendem a se formar priorizando aqueles nós que já possuem muitas conexões, o que é uma consequência, também, do fato de redes sociais não seguirem uma curva normal (como as redes aleatórias do modelo de Erdős e Rényi, por exemplo) mas sim uma *power-law*.

### 2.2.3.3 Redes do Mundo Pequeno e os Seis Graus de Separação

MILGRAM (1967) investigou o mito urbano de que entre duas pessoas existem seis passos de separação. O experimento realizado pelo referido autor era composto pela escolha aleatória de duas pessoas  $a$  e  $b$  e pela determinação da quantidade de pessoas que uma carta enviada por  $a$  passaria até que fosse recebida por  $b$ . Obviamente,  $a$  e  $b$  não poderiam ser pessoas que se conhecessem. Ao final do experimento, o autor observou que aproximadamente 1/3 das cartas chegaram ao destino em uma média de seis passos. Devido a esse trabalho, o fato ficou conhecido como “fenômeno do mundo pequeno” ou “seis graus de separação”.

O paradoxo do mundo pequeno acabou por impulsionar WATTS; STROGATZ (1998) a estudar as suas causas não somente em redes sociais, mas também em outros tipos de redes do mundo real. A partir dessa investigação, os autores concluíram que redes do mundo pequeno possuem duas características em comum que é o que tornam essas redes, de fato, “pequenas”. Primeiramente, essas redes são altamente agregadas, i.e. possuem um alto coeficiente de agrupamento e, por outro lado, são redes que possuem uma média de menor caminho regular entre nós. Assim, de acordo com os referidos autores, para se discernir se uma rede qualquer é ou não uma rede do mundo pequeno (e que conseqüentemente se a mesma se adéqua ao modelo conhecido como Watts-Strogatz), basta comparar essas propriedades com as mesmas

<b>Rede Complexa</b>	$D_{real}$	$D_{random}$	$C_{real}$	$C_{random}$
<b>Atores de Filmes</b>	3.65	2.99	0.79	0.00027
<b>Energia Elétrica</b>	18.7	12.4	0.080	0.005
<b>Rede Neural</b>	2.65	2.25	0.28	0.05

**Tabela 2.1:** Exemplo de detecção de redes do mundo pequeno (WATTS; STROGATZ, 1998).

propriedades de uma rede aleatória de mesmo tamanho da rede que se está investigando.

A Tabela 2.1 descreve o exemplo apresentado em WATTS; STROGATZ (1998) para demonstrar essa comparação. Esse exemplo compara três redes complexas reais com redes aleatórias de mesmo tamanho. Essas redes reais são nomeadas como (1) rede de atores de filmes, (2) rede de energia elétrica e (3) rede neuronal. Nesta tabela, pode-se observar que as três redes podem ser consideradas como redes do mundo pequeno, uma vez que todas possuem um coeficiente de agregação agrupamento muito alto ( $C_{real} \gg C_{random}$ ) e uma média de menor caminho regular entre os nós (i.e.  $D_{real} \sim D_{random}$ ).

#### 2.2.3.4 Redes de Livre Escala e Conexão Preferencial

O modelo de Erdős e Rényi e o modelo Watts-Strogatz explicam algumas propriedades das redes complexas encontradas no mundo real. Entretanto, esses modelos não levam em consideração o fato de que, em redes do mundo real, novos nós são frequentemente adicionados, tornando-as dinâmicas, que é exatamente o que ocorre com as redes sociais, por exemplo.

Os referidos modelos propunham, portanto, formas de geração de grafos em que a quantidade de nós era definida *a priori*, fixando o número de nós e considerando apenas a forma com que esses nós se conectavam. BARABASI; ALBERT (1999), por outro lado, observaram que esse modelo de construção não representa a realidade e, em seus estudos, mostraram que algumas redes do mundo real possuem dois mecanismos específicos de crescimento: (1) redes se expandem continuamente com a adição de novos nós e (2) os novos nós são conectados, prioritariamente, com nós que já possuem um grande número de conexões. Esses dois mecanismos foram definidos como “redes de livre escala” (ou “redes sem escala”) e conexão preferencial, respectivamente.

Redes de livre escala são redes em que nós com um alto grau (i.e. grande número de conexões) tendem a se conectar. A distribuição dos graus de uma rede de livre escala segue uma *power-law*, o que acaba por classificar as redes de livre escala como redes *power-law*.

O segundo mecanismo, denominado “conexão preferencial”, é o termo dado ao fenômeno de que ao serem introduzidos em uma rede, os nós possuem uma probabilidade maior de se conectar com outros nós que já possuem um alto grau de conexão (BARABASI; ALBERT, 1999). Esse fenômeno estende o trabalho de WATTS; STROGATZ (1998) que explicam que as redes do mundo pequeno possuem alguns poucos nós com um alto grau de coeficiente de agregação (que são, exatamente, os *hubs*). O fenômeno da conexão preferencial é conhecido também como “Efeito Mateus” (MERTON, 1968).

Com a definição das redes de livre escala e do efeito de Mateus, não somente BARABASI; ALBERT (1999) como também WATTS (2004a,b) e outros autores (NEWMAN, 2003b; LICHTENWALTER; LUSSIER; CHAWLA, 2010; NOULAS et al., 2011; CHO; MYERS; LESKOVEC, 2011; BRILHANTE et al., 2011; LEE et al., 2012) demonstraram que redes complexas do mundo real respeitam determinada dinâmica de estruturação e que essas redes são o que BARABASI; ALBERT (1999) denominam como “redes auto-organizadas”. Ou seja, as redes não crescem de forma aleatória, mas sim respeitando o fato de que os novos nós são conectados, frequentemente, a outros nós que já possuem muitas conexões.

## 2.3 Inferência de Modelos Estatísticos

Existem metodologias de inferência estatística usadas para definir a distribuição estatística que melhor se adequa a um conjunto de dados. A Seção 2.3.1 explica uma dessas metodologias, enquanto a Seção 2.3.2 apresenta um exemplo de sua aplicação.

### 2.3.1 Metodologia de Inferência Estatística

A distribuição do grau de redes sociais são comumente modeladas por meio de uma *power-law*. Para verificar a adequação desse modelo, foi usado um método de máxima verossimilhança, onde os parâmetros  $x_{min}$  e  $\alpha$  foram estimados. O teste foi realizado tal qual disposto em (CLAUSET; SHALIZI; NEWMAN, 2009), em que um método de mínimos quadrados foi utilizado e o resultado obtido foi comparado com a aproximação de outras distribuições por meio de um método de razão entre duas verossimilhanças. Para execução do teste, foi usado o pacote *powerLaw* (GILLESPIE, 2015) de R (HORNIK, 2016), uma linguagem de programação e um ambiente de desenvolvimento para cálculos estatísticos.

Primeiramente, como o interesse era descobrir se a distribuição *power-law* se ajustava bem a um conjunto de dados, CLAUSET; SHALIZI; NEWMAN (2009) propuseram utilizar um método de mínimos quadrados para se estimar os parâmetros dessa distribuição e calcular o ajuste entre a referida distribuição e o conjunto de dados. O método definido retorna um *p-value* que, baseado em um dado nível de significância, auxilia na interpretação de se a função de *power-law* se aproxima bem ou não dos dados.

Sabe-se, entretanto, que métodos dos mínimos quadrados podem produzir estimativas significativamente imprecisas dos parâmetros da distribuição *power-law*. Por exemplo, mesmo que uma distribuição *power-law* venha a se ajustar bem a um conjunto de dados, é possível ainda que outra distribuição que tenha características semelhantes a uma *power-law*, como uma exponencial ou uma log-normal, por exemplo, se ajuste tão bem ou melhor do que a *power-law*. Diante disso, CLAUSET; SHALIZI; NEWMAN (2009) propuseram, ainda, que outro teste de probabilidade auxiliasse na escolha de uma distribuição em detrimento de outra. Este teste, por sua vez, é um teste da razão de verossimilhança que auxilia a decidir quantitativamente sobre qual é a distribuição que melhor se ajusta a um conjunto de dados. Nesse teste, dois modelos

são avaliados e seus resultados são comparados por meio do cálculo do logaritmo  $\mathcal{LR}$  da razão entre as duas verossimilhanças.  $\mathcal{LR}$  é positivo ou negativo dependendo de qual distribuição melhor se ajusta. Em caso de ambas se ajustarem igualmente, o resultado é zero.

Entretanto, dependendo das flutuações estatísticas, o resultado de  $\mathcal{LR}$  pode não indicar definitivamente qual modelo melhor se ajusta ao conjunto de dados. Assim, para fazer um julgamento quantitativo, os mesmos autores sugerem que se tome conhecimento sobre as proporções dessas flutuações, ou seja, deve ser calculado o desvio padrão  $\sigma$  de  $\mathcal{LR}$ .

Para estimar o valor de  $\sigma$ , pode-se utilizar o Teste de Vuong VUONG (1989). O teste de Vuong retorna um  $p$  – *value* que indica se o valor de  $\mathcal{LR}$  é estatisticamente significativo: se  $p$  – *value* é pequeno (i.e. menor do que o nível de significância), então é improvável que o valor retornado por  $\mathcal{LR}$  seja resultante de flutuações, o que indica que o modelo resultante do teste da razão de verossimilhança é, de fato, o que melhor se ajusta aos dados. Se  $p$  – *value* é alto, por outro lado (i.e. maior do que o nível de significância), o sinal de  $\mathcal{LR}$  não é um bom indicador e o teste não favorece um modelo sobre outro.

Em linhas gerais, essa metodologia pode ser resumida como se segue:

1. Utilize um método de máxima verossimilhança para estimar  $x_{min}$  e  $\alpha$  de *power-law*;
2. Calcule o grau do ajuste entre a função empírica (i.e. conjunto de dados) e a função teórica (i.e. *power-law*). Se o  $p$  – *value* resultante for maior do que o nível de significância, então a distribuição de *power-law* pode ser considerada como uma boa escolha para modelagem dos dados, caso contrário, deve ser rejeitada;
3. Compare o grau de ajuste entre o modelo de *power-law* com outros modelos alternativos por meio do teste da razão de verossimilhança. Para cada modelo alternativo, se a razão da probabilidade é significativamente diferente de zero, então o sinal do logaritmo dessa razão (i.e.  $\mathcal{LR}$ ) indica que a distribuição alternativa se aproxima melhor aos dados do que a *power-law*;
4. Avalie o  $p$  – *value* resultante da comparação entre a função *power-law* e outro modelo alternativo, caso seu valor seja menor ou igual ao nível de significância, então o sinal de  $\mathcal{LR}$  é um bom indicador de qual distribuição melhor se aproxima do conjunto de dados. Caso contrário, nenhuma conclusão pode ser tomada;

Uma das vantagens da abordagem proposta por CLAUSET; SHALIZI; NEWMAN (2009) é que o teste não indica somente o modelo que melhor se ajusta a um conjunto de dados mas indica também quando o conjunto de dados não é suficiente para que se tome uma decisão adequada. É por essa e outras razões que diversos autores têm feito uso dessa abordagem no que diz respeito a teste de probabilidade (CLARKE, 2007; BRZEZIŃSKI, 2014).

Base de Dados	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	$\mathcal{LR}_{pl-exp}$	$p_{pl-exp}$	Resultado
<b>Pássaros</b>	<b>0.55</b>	-0.850	0.40	1.87	<b>0.06</b>	Power-law
<b>HTTP</b>	0.00	1.77	<b>0.08</b>	11.80	<b>0.00</b>	Nenhuma

Tabela 2.2: Exemplo de Teste de Vuong.

### 2.3.2 Exemplo de Seleção de Modelo Estatístico

Esta seção apresenta um simples exemplo para que os conceitos da Seção 2.3.1 sejam melhor compreendidos. Este exemplo foi apresentado no trabalho de CLAUSET; SHALIZI; NEWMAN (2009) que descreve as seguintes bases de dados:

- **Pássaros:** base de dados contendo o número de vezes em que diferentes espécies de pássaros foram observados na América do Norte em 2003;
- **HTTP:** base de dados contendo a quantidade de arquivos web (em bytes) transmitidos pela Internet em junho de 1996 (PARK; WILLINGER, 2000);

Para se discernir sobre a aproximação que melhor se ajusta a cada uma das bases de dados, foi calculado primeiramente o grau de ajuste de uma *power-law* com os referidos conjuntos de dados. Feito isto, esta aproximação foi comparada com a aproximação de uma distribuição log-normal e uma exponencial com o objetivo definir qual distribuição melhor se aproxima da base de dados em análise. Para isto, o seguinte teste de hipóteses foi considerado:

$H_0$ : Ambas distribuições são igualmente distantes da distribuição verdadeira

$H_1$  : Uma das distribuições de teste é próxima da distribuição verdadeira.

A Tabela 2.2 apresenta o resultado do teste de probabilidade. Considere, para este exemplo, um nível de significância igual a 90%. No primeiro caso, para a base pássaros, nota-se que a distribuição que melhor se aproxima dos dados é a *power-law*, uma vez que além do ajuste dessa função ter resultado em um *p-value* aceitável, pode-se perceber que ao se comparar essa aproximação com a aproximação da log-normal e da exponencial, o teste de Vuong reforçou o primeiro resultado. Ou seja, embora  $\mathcal{LR}_{pl-ln}$  tenha resultado em um valor negativo, que indicaria que a log-normal teria uma aproximação melhor, o valor de  $p_{pl-ln}$  foi maior do que o nível de significância, indicando que o sinal de  $\mathcal{LR}_{pl-ln}$  não é confiável. Por outro lado, enquanto que  $\mathcal{LR}_{pl-exp}$  reforça que a aproximação de uma *power-law* se ajusta melhor do que a aproximação de uma exponencial, o valor de  $p_{pl-exp}$  indica que, de fato, o sinal de  $\mathcal{LR}_{pl-exp}$  é confiável. Para a base HTTP, pode-se notar que não foi possível definir qual das distribuições analisadas poderia ser a que melhor se ajusta à base de dados.

## 2.4 Considerações Finais

Neste capítulo foi apresentada a fundamentação teórica necessária para compreensão do trabalho proposto neste documento. Na Seção 2.1 foram apresentadas as principais definições

---

bem como um breve histórico da área de redes sociais online, incluindo a evolução dos serviços de localização geográficas nesse tipo de rede. A Seção 2.2, por sua vez, listou conceitos da teoria dos grafos, assim como as propriedades e padrões relacionados a esse tipo de estrutura. Finalmente, a Seção 2.3 descreveu o teste de inferência para seleção de modelos estatísticos que foi utilizado nesta tese, e seus resultados serão mostrados no Capítulo 4 e no Capítulo 5.

## 3

### Trabalhos Relacionados

Embora redes sociais seja um tópico previamente definido em outras áreas, a implementação desse conceito na web tem demonstrado dois fenômenos importantes desde o início do presente século. O primeiro, diz respeito à popularização de sites de rede sociais, como o Facebook, por exemplo. Atualmente, este tipo de serviço é considerado o mais popular da web (AMAZON, 2016). O segundo fenômeno é a grande quantidade de estudos relacionados ao tema. Pesquisadores das mais diversas áreas têm se preocupado em definir teorias, ferramentas e processos com o objetivo de compreender os relacionamentos e a própria estrutura da rede (HOPPE; REINELT, 2010).

Na mesma medida em que cresce o número de usuários de redes sociais online, aumenta a diversidade e sofisticação desse tipo de serviço (GROSS; ACQUISTI, 2005). Conseqüentemente, os mais variados tipos de usuários têm feito uso desse tipo de ferramenta online com diferentes objetivos, seja para conhecer pessoas (e.g. Facebook), seja em busca de contatos profissionais (e.g. LinkedIn) ou, simplesmente, para compartilhar vídeos (e.g. YouTube<sup>1</sup>).

Nesse contexto, surgem também as redes sociais online equipadas com serviços de localização. Esse tipo de rede se popularizou rapidamente, atraindo um número sempre crescente de usuários interessados em compartilhar os locais visitados. Não muito diferente, portanto, a comunidade acadêmica passou a produzir uma série de estudos sobre a análise de dados capturados nesse tipo de rede. Este capítulo apresenta alguns dos trabalhos que fundamentaram esta tese e está organizado de acordo com a taxonomia apresentada na Seção 3.1. Essa taxonomia descreve diferentes categorias de trabalhos relacionados. Essas categorias foram definidas considerando-se as contribuições que cada um dos trabalhos classificados trouxeram para esta tese bem como considerando-se as contribuições que esta tese traz para cada uma das áreas apresentadas.

#### 3.1 Introdução

O trabalho de MISLOVE et al. (2007) é conhecido como um dos primeiros estudos da área de redes sociais online que fez uso de uma grande quantidade de dados. Os autores

---

<sup>1</sup><https://www.youtube.com/>

examinaram os referidos dados e a partir deles construíram grafos. Esses autores confirmaram os resultados já apresentados por MILGRAM (1967) e WATTS; STROGATZ (1998), indicando que as interações sociais entre indivíduos podem ser modeladas por uma *power-law*. Ou seja, enquanto que alguns poucos indivíduos possuem uma grande quantidade de laços sociais, a maioria possui poucos laços sociais.

Na área de redes sociais equipadas com serviços de localização, o trabalho de LI; CHEN (2009) foi pioneiro. Além de terem demonstrado que em redes do tipo LBSN a distribuição do número de amigos por usuários também é governada por uma *power-law*, esse trabalho extraiu fatores relativos à mobilidade dos usuários que foram classificados em quatro tipos diferentes: (1) usuários caseiros, (2) usuários em férias, (3) usuários trabalhadores e, por fim, (4) outros usuários, que são todos aqueles que não foram classificados nas três primeiras classes.

Assim como o seminal trabalho de MISLOVE et al. (2007) foi proeminente para que outros estudos fossem desenvolvidos na área de redes sociais online como um todo, o trabalho de LI; CHEN (2009) influenciou pesquisadores a investigar também os dados espaço-temporais de redes equipadas com serviços de localização. Em (LINDQVIST et al., 2011), uma pesquisa qualitativa e quantitativa é detalhada e indica as possíveis razões pelas quais as pessoas compartilham os locais visitados em redes sociais online.

Portanto, são muitas as possibilidades de análise de dados de redes sociais online, incluindo as redes equipadas com serviços de localização. Diante da grande quantidade de trabalhos relacionados bem como das diferentes perspectivas com que esses trabalhos influenciaram o desenvolvimento desta tese, foi definida uma taxonomia para simplificar a apresentação e descrição dos mesmos. Esta taxonomia é ilustrada na Figura 3.1 e foi usada para organizar o restante deste capítulo. É importante observar que determinados trabalhos correlatos podem sobrepor mais de uma categoria por pertencerem a mais de uma classe da taxonomia proposta.

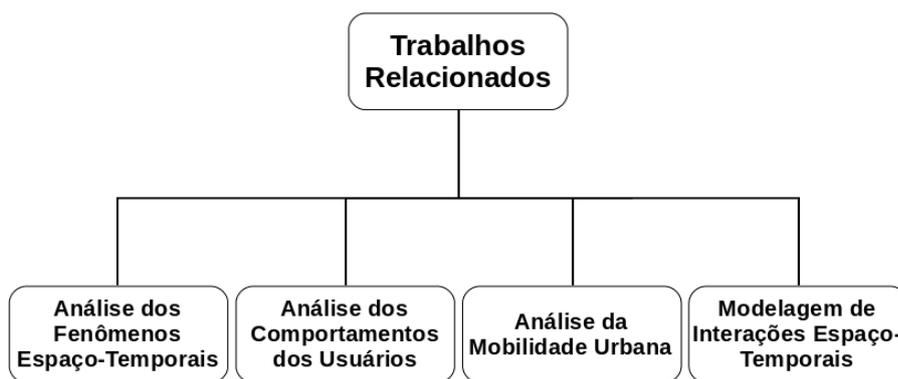


Figura 3.1: Categorias de classificação de trabalhos relacionados

### 3.2 Análise de Fenômenos Espaço-Temporais

Já é consenso o fato de que as relações sociais são governadas por meio de uma *power-law* (WATTS; STROGATZ, 1998; LI; CHEN, 2009; MOYANO; THOMAE; FRÍAS-MARTÍNEZ,

2012). No que diz respeito às relações espaciais e espaço-temporais dos indivíduos, diferentes resultados têm sido obtidos dependendo do ponto de vista de que essas relações são estudadas.

Tanto em (NOULAS et al., 2011), como em (BACKSTROM; SUN; MARLOW, 2010) e (MOYANO; THOMAE; FRÍAS-MARTÍNEZ, 2012), por exemplo, os autores mostraram que a distribuição do deslocamento de usuários de redes sociais equipadas com serviços de localização segue uma *power-law*. Em (SCELLATO; MASCOLO, 2011), por outro lado, os autores indicaram que o número de *check-ins* publicados por usuários de redes sociais é melhor distribuído por uma log-normal. Essa descoberta pode ser motivada por restrições físicas, uma vez que usuários são impossibilitados de visitar uma grande quantidade de locais enquanto que laços sociais podem ser feitos virtualmente sem qualquer custo físico. Estes resultados mostram que o engajamento entre usuários de redes sociais é diferente quando são requeridas ações no mundo real. Esse resultado é, inclusive, bastante semelhante ao apresentado em (SESHADRI et al., 2008).

Embora distribuições log-normais e *power-law* sejam comumente confundidas (MIT-ZENMACHER, 2004), é de fundamental importância que se saiba por qual distribuição um conjunto de dados é governado. Sabendo-se a distribuição que melhor se aproxima de um conjunto de dados, pode-se construir um modelo capaz de representar os dados reais. Além disso, diferentes modelos teóricos possuem diferentes características: um grafo com conexões governadas por uma *power-law*, por exemplo, possui nós conectados uns aos outros obedecendo o fenômeno de conexão preferencial onde os nós tendem a se conectar com nós de mais alto grau, como já demonstrado por BARABASI; ALBERT (1999); já uma rede governada por uma log-normal indica que além de os nós possuírem a tendência de se conectar preferencialmente, esses nós também podem se conectar aleatoriamente (JACKSON; ROGERS, 2007). Ou seja, em grafos modelados por uma log-normal, a tendência de conexão preferencial não é tão latente como no caso de um grafo governado por uma *power-law* (YIP; SHADBOLT; WEBBER, 2012).

Em (KYLASA; KOLLIAS; GRAMA, 2015), é mostrado que usuários que publicam muitos *check-ins* são usuários que possuem um alto grau de conectividade social, entretanto, o inverso não é verdadeiro, ou seja: uma conectividade social alta não implica em uma grande quantidade de *check-ins* compartilhados. Como consequência disso, usuários que realizam muitos *check-ins* possuem uma forte tendência de se agrupar com outros usuários, ao contrário daqueles que publicam poucos *check-ins*.

Ainda de acordo com KYLASA; KOLLIAS; GRAMA (2015), quando se deseja investigar de que modo a informação flui em uma rede social, essa investigação pode ocorrer por meio da análise dos fenômenos espaço-temporais, uma vez que estes se sobrepõem aos fenômenos meramente sociais, influenciando-os diretamente. Em (CHO; MYERS; LESKOVEC, 2011), por outro lado, é indicado que a dinâmica dos movimentos das pessoas possui forte implicação dos seus laços sociais. Por exemplo, movimentos pequenos, de característica rotineira, não são afetados pelas relações sociais dos indivíduos, ao contrários de movimentos mais longos, como uma viagem, por exemplo, que normalmente pode ser explicado por razões sociais.

### 3.3 Análise do Comportamento de Usuários

VASCONCELOS et al. (2012) fazem uso das três funcionalidades públicas do Foursquare - *tips*, *done*s e *to-dos* - para classificar os usuários em diferentes tipos. Basicamente, as “*tips*” são dicas públicas que usuários fazem em seus *check-ins* e que podem ser visualizadas por outros usuários. Uma *tip* pode ser marcada por outro usuário como “*done*” (realizada), indicando que o referido usuário concorda com aquela dica e pode, também, ser adicionada a uma lista de afazeres do usuário, i.e. a lista de “*to-dos*”. A partir da análise dessas três funcionalidades bem como do comportamento dos usuários, quatro tipos de usuários foram definidos: (1) usuários ocasionais (i.e. que postam *tips* em poucos lugares e recebem apenas poucos *done*s e *to-dos*); (2) usuários ativos (i.e. que tendem a obter um *feedback* maior em relação às *tips* postadas); (3) usuários caracterizados por receberem um alto *feedback* (i.e. postam *tips* em grande quantidade e recebem uma grande quantidade de *done*s e *to-dos* como resposta) e, por fim, (4) usuários *spammers*, caracterizados por uma grande quantidade de *tips* contendo links e postadas em diferentes locais. A Figura 3.2(a) mostra as palavras mais utilizadas nas *tips* de usuários classificados como *spammers* e a Figura 3.2(b) exibe aquelas que são mais publicadas por usuários considerados *não-spammers*.



(a) Usuários rotulados como spammers

(b) Usuários rotulados como não-spammers

**Figura 3.2:** Palavras mais utilizadas por usuários do Foursquare nas *tips* (VASCONCELOS et al., 2012)

Na Figura 3.2, fica evidente como usuários considerados *spammers* comportam-se diferente de outros tipos de usuários. *Spammers* fazem uso de palavras relacionadas a anúncio de produtos ou de serviços, conforme mostrado na Figura 3.2(a) por meio das palavras *business*, *apartment*, *iphone*, *free*, *gadget*, dentre outras.

VASCONCELOS et al. (2014) mostraram que diferentemente de outros tipos de conteúdos gerados em redes sociais online, as *tips* do Foursquare são caracterizadas por uma evolução mais lenta de popularidade, ou seja, só após um longo tempo conseguem atrair a atenção de uma grande quantidade de usuários. Foi mostrado que os laços sociais dos usuários que publicam conteúdo em uma rede social como o Foursquare possuem um importante papel para que o conteúdo do referido usuário seja, de fato, popular e definem o tempo de popularidade deste conteúdo.

Assim como em (VASCONCELOS et al., 2012) e (VASCONCELOS et al., 2014), os autores de (GEORGIEV; NOULAS; MASCOLO, 2014a) também fizeram uso de dados do

Foursquare porém com um objetivo diferente: tentaram compreender quais são os fatores que influenciam as pessoas a participarem de eventos. Nesse estudo, os autores concluíram que apesar de a popularidade do evento ser um forte atrativo, os aspectos sociais são, de fato, fundamentais para que usuários realizem *check-ins* em locais que estão ocorrendo eventos, como uma conferência, por exemplo. O trabalho de JIANG; LI (2016) fez uma análise ainda mais detalhada a respeito desses fatores, considerando dados de uma rede social baseada em eventos.

Já em uma linha de pesquisa mais voltada para aspectos psicológicos, DONG; JIN; KNIJNENBURG (2015) estudaram, por sua vez, os fatores que podem levar os usuários a compartilharem informações privadas, ou sensíveis. Foi mostrado que a “*confiança*” na audiência (i.e. as pessoas que visualizam os *check-ins*) atua como um fator influenciador da decisão do usuário sobre o compartilhamento ou não de sua localização geográfica. Isso significa que é mais provável que um usuário de uma rede do tipo LBSN compartilhe um local que o mesmo visitou para um amigo ou um familiar no lugar de compartilhá-lo com um colega de trabalho, por exemplo.

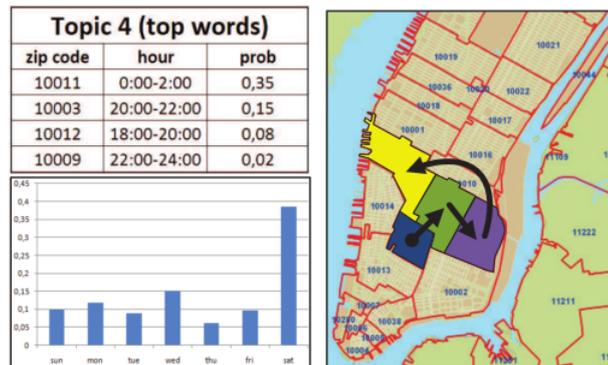
Além da confiança, DONG; JIN; KNIJNENBURG (2015) observaram que a “*privacidade*” do local também atua como um importante fator que influencia a decisão do usuário em compartilhar ou não sua localização. Ou seja, existe uma pequena probabilidade de usuários publicarem que estão em um banco ou em uma delegacia de polícia, por exemplo, o que não ocorre com restaurantes, cinemas ou shopping centers. Outro fator que também é bastante considerado por usuários antes de compartilhar sua localização é a “*adequação*”. A adequação diz respeito ao fato de que diferentes categorias de locais são, normalmente, compartilhadas com diferentes públicos. Por exemplo, é mais provável que usuários compartilhem com seus familiares de que o mesmo está em um hospital do que com seus amigos. Isso pode não ser verdade quando o local é um bar ou um restaurante, por exemplo.

Uma importante conclusão de DONG; JIN; KNIJNENBURG (2015) foi o fato de que existe uma alta tendência de um usuário compartilhar na rede social que está no mesmo local que outros amigos. De acordo com os autores, esta observação é, inclusive, consistente com o fator psicológico de que as pessoas gostam de demonstrar que são aceitas por um grupo de pessoas.

### 3.4 Análise da Mobilidade de Usuários

FERRARI et al. (2011) fizeram uso de dados geo localizados do Twitter e de um modelo probabilístico LDA para avaliar padrões de mobilidade espacial de usuários da cidade de Nova Iorque. Modelos LDA são comumente utilizados para mineração de texto e consideram cada palavra de um determinado documento como sendo a unidade básica para análise do texto. Neste caso, os autores definiram que o CEP do local (i.e. o *zip code*) seria a palavra a ser analisada pelo modelo LDA. Este modelo adaptado foi capaz de identificar padrões de mobilidade considerando-se tanto a dimensão espacial quanto a dimensão temporal. O modelo revelou também características bem conhecidas da movimentação urbana. A Figura 3.3, por exemplo,

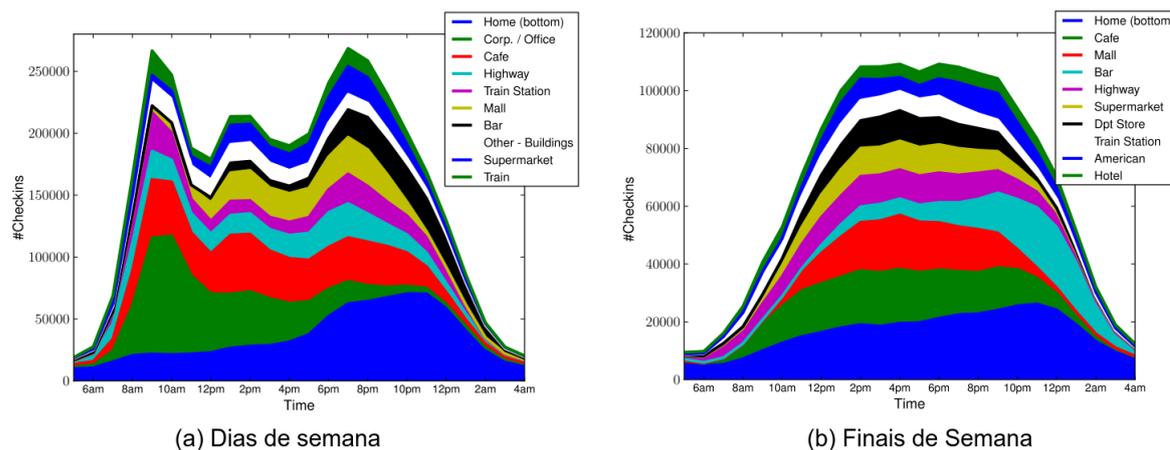
ilustra um dos padrões capturados pelos autores, que consiste em um padrão de movimentação da vida noturna em Nova Iorque. O lado esquerdo apresenta a probabilidade de um referido local (i.e. *zip code*) ser frequentado em um dado instante de tempo (i.e. *hour*). Esse padrão é exibido no mapa do lado direito, ilustrando um padrão sequencial da movimentação dos usuários tal qual relatado nas palavras do lado esquerdo da figura. O gráfico da Figura 3.3 mostra também a distribuição dos locais de acordo com os dias da semana.



**Figura 3.3:** Comportamento rotineiro de usuários em um sábado a noite em Nova Iorque (FERRARI et al., 2011)

NOULAS et al. (2011) estudaram a mobilidade dos usuários por meio das categorias dos locais em que os mesmos fizeram *check-in*. Eles descobriram que as atividades de usuários de redes sociais baseadas em localização variam no decorrer da semana e esses *check-ins* ilustram, de fato, padrões muito semelhantes às atividades do dia a dia de uma pessoa. Por exemplo, a Figura 3.4(a) relaciona o número de *check-ins* com as categorias de locais do Foursquare para os dias da semana, a Figura 3.4(b), por sua vez, faz o mesmo para os finais de semana. Pode-se observar que as duas curvas apresentam padrões consideravelmente diferentes. Nos dias de semana pode-se observar três categorias mais utilizadas: enquanto que pela manhã as pessoas estão no trabalho, na hora do almoço e entre 18 e 20 horas, as pessoas estão em casa ou em shoppings centers ou bares. Já nos fins de semana, a curva indica que os locais mais frequentados são locais com atividades de lazer predominantes.

O impacto da distância espacial nas relações sociais dos usuários foi estudado (BACKSTROM; SUN; MARLOW, 2010; SCELLATO et al., 2011; MOYANO; THOMAE; FRÍAS-MARTÍNEZ, 2012). BACKSTROM; SUN; MARLOW (2010) fizeram uso de dados do Facebook bem como de informações de localização fornecidas pelos próprios usuários para investigar a relação entre os laços sociais dos usuários e a distância entre eles. Em um estudo semelhante mas fazendo uso de dados da rede telefônica, MOYANO; THOMAE; FRÍAS-MARTÍNEZ (2012) analisaram dados espaço-temporais com o objetivo de avaliar até que ponto a mobilidade do usuário influencia na formação da rede social do indivíduo. Os resultados indicaram que existe uma dependência entre a estrutura social da rede e a distribuição espacial dos usuários. Já em (SCELLATO et al., 2011), os autores mostraram que além da distância, outras propriedades podem influenciar as interações sociais dos usuários.

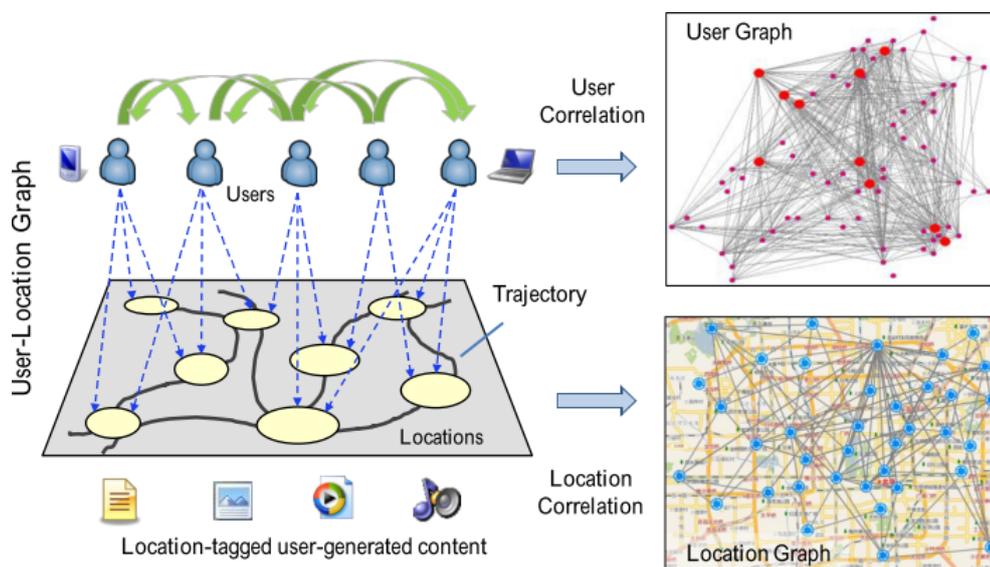


**Figura 3.4:** As 10 categorias mais populares distribuídas em (a) dias de semana e (b) finais de semana (NOULAS et al., 2011)

RIEDERER et al. (2015) usaram fotografias geo-localizadas para compreender a mobilidade de usuários de redes sociais e detectar padrões de mobilidade. Além de terem reproduzido alguns dos resultados já descritos neste capítulo (como o fato de que o deslocamento de um indivíduo pode ser modelado por uma *power-law*), os autores também revelaram que existem diferenças significativas na mobilidade de usuários quando informações contextuais do usuário, como o gênero ou sua etnia, são consideradas.

### 3.5 Modelagem de Fenômenos Espaço-Temporais

Em redes sociais online equipadas com serviços de localização, a informação espacial permite que novas relações sejam estabelecidas. Baseado nas informações sociais e espaço-temporais, diferentes tipos de grafos podem ser construídos. BAO et al. (2013) elencaram alguns deles conforme mostra a Figura 3.5.



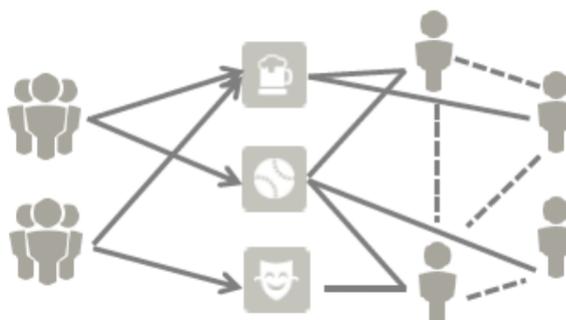
**Figura 3.5:** Tipos de grafos construídos a partir de redes do tipo LBSN (BAO et al., 2013).

Em (ALLAMANIS; SCELLATO; MASCOLO, 2012), os autores definiram um tipo específico de grafo para representar as relações entre usuários de redes do tipo LBSN. Cada usuário de uma rede social desse tipo é representado como um nó do referido grafo, enquanto as conexões entre esses nós representam as interações sociais entre os usuários. Esse modelo replica os grafos a medida em que as interações sociais evoluem com o tempo. A informação espacial capturada a partir dos *check-ins* dos usuários é representada por meio de uma medida de distância entre os usuários conectados e essa medida é um atributo da conexão entre cada par de usuário. Baseado neste modelo, os autores avaliaram de que modo essas interações espaço-temporais entre usuários ocorrem com o passar do tempo.

Em (BORA et al., 2013), dados espaço-temporais capturados no Twitter foram usados para compreender as relações existentes entre regiões, bairros e territórios dominados por gangues da cidade de Los Angeles. Para isso, um modelo foi definido onde cada nó representa uma rua dominada por uma gangue e as conexões entre os nós representam as interações ocorridas entre pessoas que fazem parte de diferentes gangues. Por exemplo, se uma pessoa de uma gangue “X” visita uma rua dominada por uma gangue “Y”, a rua de X é conectada à rua de Y que foi visitada. O modelo proposto foi utilizado em conjunto com técnicas de aprendizagem de máquina com o objetivo de prever possíveis interações entre gangues rivais.

Para identificar preferências de usuários com relação aos eventos em que os mesmos participam, GEORGIEV; NOULAS; MASCOLO (2014a) definiram um modelo que combina informações sociais e espaciais dos usuários, bem como a categoria dos locais e o tipo dos eventos em que os mesmos participam. A Figura 3.6 ilustra um exemplo do grafo proposto, o qual além de ser composto por nós que representam usuários, possuem nós que também representam eventos e locais. As conexões também representam conceitos diferentes: enquanto que as conexões com linha tracejada não são direcionadas e descrevem relações sociais entre usuários, as conexões entre eventos e os tipos dos locais são contínuas e direcionadas e as conexões entre usuários e os referidos tipos de locais são, também, contínuas, porém não direcionadas.

De acordo com os autores, além de o referido modelo auxiliar na recomendação e previsão de eventos para determinados usuários, o mesmo auxilia também na definição de serviços e aplicações baseadas em localização.



**Figura 3.6:** Exemplo de grafo tal qual definido em (GEORGIEV; NOULAS; MASCOLO, 2014a).

O sistema de recomendação proposto em (YUAN; CONG; SUN, 2014) foi baseado

em um modelo de grafos definido com o objetivo de representar as relações sociais e espaço-temporais de usuários de redes sociais online. O referido modelo permite a criação de grafos compostos por três tipos de nós diferentes: *nó usuário*, *nó POI* (i.e. local) e *nó sessão*. Além dos diferentes tipos de nós, o modelo também define dois tipos diferentes de conexões: *conexão check-in* e *conexão POI*. Uma sessão representa um *slot* de tempo (e.g. uma hora do dia). Neste modelo, cada *nó usuário* se conecta com um conjunto de sessões de um usuário e cada *nó sessão* conectado a um usuário é conectado a um *nó POI* se o usuário daquela sessão visitou aquele POI no *slot* identificado pela referida sessão.

De acordo com o modelo definido por YUAN; CONG; SUN (2014), um par de conexões (entre um *nó usuário* e um *nó sessão* e entre um *nó sessão* e um *nó POI*) formam uma *conexão check-in*, que representa, de fato, um *check-in* do usuário. Por outro lado, dois *nós POIs* são conectados através de uma *conexão POI* se eles são localizados geograficamente próximos um do outro. Dessa forma, o modelo definido pelos autores consegue incorporar tanto a informação temporal como a geográfica de cada *check-in*.

### 3.6 Considerações Finais

Diante dos trabalhos apresentados neste capítulo, pode-se notar que a análise das interações sociais e espaço-temporais sob a perspectiva de encontros entre usuários é, ainda, uma novidade na área de redes sociais online. Assim, ainda se faz necessário investigar problemáticas relacionadas a cada uma das categorias da taxonomia apresentada na Figura 3.1. Essas problemáticas são resumidas a seguir:

- A **análise dos fenômenos espaço-temporais** da Seção 3.2 mostrou que embora relações sociais entre usuários sejam governadas por uma *power-law*, ainda não há consenso sobre como fenômenos espaço-temporais são modelados. Enquanto alguns autores indicam que o deslocamento de usuários pode ser explicado por uma *power-law*, outros descreveram que a publicação de *check-ins* é melhor aproximada por uma log-normal.
- Para **comportamento dos usuários**, pode-se observar que os laços sociais dos usuários atuam como um importante fator na publicação de *check-ins*. Além disso, de acordo com o comportamento dos usuários, diferentes autores têm proposto classificar os usuários em diferentes tipos. Entretanto, ainda não está claro como usuários se comportam quando desejam publicar que estão no mesmo local do que outros usuários.
- Os trabalhos relacionados à **mobilidade dos usuários** indicam que muitos trabalhos têm fornecido valiosos estudos utilizando-se *check-ins* que podem ser comparados com estudos que fizeram uso de outros dados gerados a partir de fontes alternativas. Entretanto, até o momento de escrita dessa tese não se tinha conhecimento de estudos

relacionados à área de mobilidade de usuários que fizessem uso de encontros de usuários publicados em redes sociais online.

## 4

### Detecção de Encontros

Quando um usuário de uma rede social online realiza *check-in* em um local, ele está declarando que está visitando aquele local em um dado instante de tempo. Ao mesmo tempo em que esse usuário realiza o *check-in*, outros usuários - amigos ou não do primeiro - podem, por sua vez, visitar o mesmo local no mesmo momento em que o primeiro o fez, produzindo interações espaço-temporais. Essas interações são aqui denominadas como encontros espaço-temporais entre usuários ou, simplesmente, encontros entre usuários.

A análise dos encontros entre usuários e sua caracterização oferece novos caminhos para se modelar fenômenos relacionados ao comportamento humano, aos padrões de movimento e a mobilidade de usuários. O objetivo deste capítulo é avaliar como esses encontros acontecem sob diferentes perspectivas. Para que estes encontros sejam avaliados, portanto, é necessário um algoritmo que seja capaz de detectá-los a partir de um conjunto de *check-ins* e de acordo com parâmetros pré-definidos. Diante disso, a análise de encontros apresentada nesse capítulo é introduzida com a apresentação do algoritmo de detecção de encontros.

Esta análise é realizada, primeiramente, por meio da investigação da distribuição de encontros por usuários que os publicaram e por locais em que foram publicados. Ou seja, foi feita uma descrição da probabilidade de um usuário publicar o evento de encontro com outro usuário em um local. Além disso, este capítulo também apresenta resultados da análise desses encontros em relação ao tempo e os possíveis padrões que podem emergir a partir destes resultados.

Este capítulo está organizado como se segue. A Seção 4.1 lista as definições básicas desta tese. A Seção 4.2 apresenta o algoritmo proposto para detectar encontros. Na Seção 4.3, é detalhada a metodologia utilizada para análise das distribuições estatísticas realizadas. Na Seção 4.4, as bases de dados utilizadas no decorrer da pesquisa são descritas. Finalmente, a Seção 4.5 exhibe os resultados da detecção de encontros nas bases de dados utilizadas.

#### 4.1 Definições Básicas

A funcionalidade de *check-in*, presente na maioria das redes sociais online mais populares, tem criado um ambiente que permite usuários compartilharem suas localizações em um determinado instante de tempo. Esta funcionalidade acaba por criar uma conexão entre o mundo

físico (ou seja, os movimentos do usuário no globo terrestre) e o mundo virtual (o que o usuário publica a respeito dos seus movimentos nas redes sociais online).

Um encontro entre usuários ocorre quando dois ou mais usuários estão em um mesmo local ao mesmo tempo e por razões diversas e desconhecidas decidem publicar este encontro nas redes sociais online através de seus *check-ins*. Estes encontros são caracterizados pela publicação de informações espaciais (já que estão frequentando o mesmo local), temporais (uma vez que esse encontro ocorre em um instante de tempo aproximado) e, também, informações sociais. Este evento, i.e. o encontro, é o objeto de estudo desta tese e é definido pela Definição 1.

**Definição 1. Encontro:** Os usuários  $u_a$  e  $u_b$  de um conjunto de  $N$  usuários são participantes de um encontro se eles publicaram *check-ins* no mesmo local  $p$  e respeitando um intervalo de tempo de encontro (i.e. intervalo de encontro)  $\Delta t$ .

Como redes sociais online são caracterizadas por apresentarem os laços de amizade entre os usuários, a definição de encontro de usuários pode ser especializada em dois tipos adicionais de encontros: (1) *encontros sociais* (Definição 2) e (2) *encontros não sociais* (Definição 3).

**Definição 2. Encontros sociais:** Os usuários  $u_a$  e  $u_b$  de um conjunto de  $N$  usuários são participantes de um *encontro social* se eles estão envolvidos em um encontro e se possuem laços sociais publicados na rede social;

**Definição 3. Encontros não sociais:** Os usuários  $u_a$  e  $u_b$  de um conjunto de  $N$  usuários são participantes de um *encontro não social* se eles estão envolvidos em um encontro e não possuem laços sociais publicados na rede social;

A Definição 2 diz que se dois usuários possuem laços sociais expressos na rede social online e eles se encontraram em algum lugar respeitando um *intervalo de encontro* ( $\Delta t$ ), então eles são participantes de um encontro social. Por outro lado, a Definição 3 mostra que se dois usuários se encontram no mesmo local, respeitando o intervalo de encontro mas seus laços sociais não estão publicados, então eles são participantes de um encontro não social. Ou seja, mesmo que estes últimos tenham laços sociais no mundo real, se esses laços não estão expressos na rede social online, o encontro é considerado como sendo um encontro não social.

## 4.2 Algoritmo para Detecção de Encontros

A detecção de encontros é descrita no Algoritmo 1. Este algoritmo recebe os seguintes parâmetros de entrada: (1) uma lista  $Ck$  de *check-ins* agrupados por local e ordenados pelo instante de tempo, (2) um intervalo de tempo, i.e. *intervalo de encontro*,  $\Delta t$  e (3) um valor booleano ( $st$ ), que especifica se os laços sociais dos usuários devem ser considerados ou não. Ou seja, esse atributo especifica o tipo de encontro a ser detectado. A Figura 4.1 representa este algoritmo graficamente por meio de um diagrama de atividades em UML.

**Algoritmo 1** detectaEncontros ( $Ck, \Delta t, st$ )

---

```

1:  $ENC \leftarrow \{ \}$  // lista de encontros a ser retornada (inicialmente vazia)
2:  $i := 0$  // variável de incremento do PARA a seguir
3: PARA cada par ordenado  $ck_i$  e  $ck_{i+1}$  de check-ins de  $Ck$ :
4:   SE  $ck_i$  e  $ck_{i+1}$  foram publicados por usuários diferentes:
5:     SE  $ck_i$  e  $ck_{i+1}$  ocorreram no mesmo local:
6:       SE a diferença temporal entre  $ck_i$  e  $ck_{i+1}$  respeita  $\Delta t$ :
7:         SE ( $st$  é falso) OU (se  $st$  é verdadeiro E  $ck_i$  e  $ck_{i+1}$  possuem laços sociais):
8:            $ENC \leftarrow ENC + (ck_i, ck_{i+1})$ 
9:         FIM SE
10:      FIM SE
11:     FIM SE
12:   FIM SE
13:    $i := i+1$ ;
14: FIM PARA
15: RETORNE  $ENC$ ;

```

---

Como os *check-ins* (parâmetro  $Ck$ ) são ordenados pelo local e pelo instante de tempo, o Algoritmo 1 basicamente compara cada par de *check-in* com o objetivo de identificar se ambos são participantes de um encontro tal qual a Definição 1. Na linha 1 do referido algoritmo, uma lista  $ENC$  é inicializada. Cada unidade desta lista será composta por um par de *check-ins* que representarão, cada um, uma instância de encontro. Cada par de encontro, por sua vez, é processado na estrutura de repetição entre as linhas 3 e 14. Em cada iteração, o algoritmo verifica se os pares de *check-ins* foram publicados por usuários diferentes (linha 4), se ocorreram no mesmo local (linha 5) e se a diferença temporal entre ambos respeita o intervalo de encontro ( $\Delta t$ ), na linha 6. Finalmente, na linha 7, o algoritmo avalia se devem ser detectados encontros sociais ou não sociais.

Caso o parâmetro  $st$  seja falso, isto significa que os encontros a serem detectados pelo algoritmo são encontros não sociais e, portanto, não se deve considerar os laços sociais que tenham sido eventualmente publicados nas redes sociais. Caso  $st$  seja verdadeiro, o algoritmo verifica se os usuários que publicaram os *check-ins* da iteração corrente possuem laços sociais. Dependendo da circunstância (i.e. se  $st$  é falso ou se  $st$  é verdadeiro e os usuários possuem laços sociais), o par de encontro da iteração corrente é adicionado na lista de retorno (linha 8).

### 4.3 Metodologia para Análise das Distribuições Estatísticas

O Capítulo 3 mostrou que a distribuição de *power-law* é comumente utilizada na literatura como a que mais se aproxima de dados gerados a partir de interações sociais de usuários. No que diz respeito aos fenômenos espaço-temporais, o mesmo capítulo mostrou que enquanto alguns trabalhos utilizam a *power-law* para modelar esse tipo de fenômeno, outros estudos têm indicado que a função log-normal pode ser uma escolha melhor.

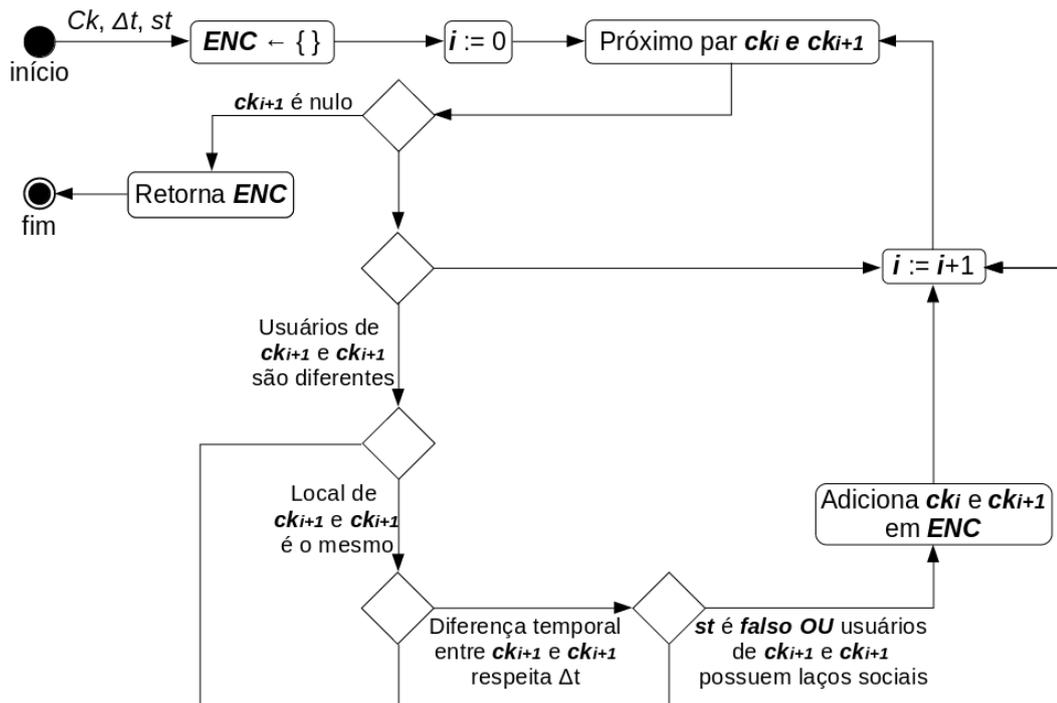


Figura 4.1: Diagrama de atividades do Algoritmo 1.

Assim, com o objetivo de investigar quais leis melhor se ajustam aos dados provenientes de encontros, o presente capítulo faz uso de uma abordagem que, primeiramente, tenta aproximar uma *power-law* aos dados de encontros. Depois disto, faz-se também uma comparação daquela aproximação com uma aproximação de outros três modelos teóricos: log-normal, exponencial e Poisson. Os modelos log-normal e exponencial foram escolhidos porque, graficamente, foram os que mais se aproximaram aos conjuntos de dados analisados no decorrer deste trabalho. O modelo de Poisson, por sua vez, foi escolhido pelo fato de que a literatura demonstra que eventos aleatórios são governados, em sua maioria, por essa função, como pode ser visto na Seção 2.2.3.2. Deste modo, nas diversas aplicações da metodologia para análise das distribuições estatísticas aqui apresentada, tem-se o objetivo de investigar, também, se as interações entre usuários e locais são eventos aleatórios ou não.

Essa primeira comparação é visual, feita por meio da análise da Função de Densidade Acumulada (FDA) para cada distribuição. Em seguida, um teste de inferência estatística, o Teste de Vuong (apresentado no Capítulo 3), foi executado com o objetivo de definir qual distribuição, de fato, se ajusta melhor aos encontros de usuários.

Em todo o teste de inferência foi escolhido um nível de significância de 90%, já na razão entre duas verossimilhanças (i.e. o Teste de Vuong), foram consideradas as seguintes hipóteses:

$H_0$ : Ambas distribuições são igualmente distantes da distribuição verdadeira

$H_1$ : Uma das distribuições de teste é próxima da distribuição verdadeira.

#### 4.4 Descrição das Bases de Dados

Nos experimentos realizados, diferentes bases de dados reais foram usadas. Estes dados foram gerados voluntariamente por usuários de redes sociais online e foram capturados por meio das APIs oficiais de cada uma dessas redes.

Seis bases de dados de quatro redes sociais distintas foram utilizadas. Das seis bases de dados, uma foi capturada a partir do Brightkite, uma do Gowalla, duas do Gowalla e, finalmente, duas do Twitter. A Tabela 4.1 apresenta os nomes adotados para identificar cada uma dessas bases de dados bem como de onde as mesmas foram obtidas. Os experimentos apresentados no presente capítulo e no Capítulo 5 fizeram uso das bases descritas na Tabela 4.1.

	<b>Nome</b>	<b>Rede</b>	<b>Fonte</b>
1	BKT_CHO	Brightkite	CHO; MYERS; LESKOVEC (2011)
2	GOW_CHO	Gowalla	CHO; MYERS; LESKOVEC (2011)
3	GOW_SCE	Gowalla	SCCELLATO; NOULAS; MASCOLO (2011)
4	4QR_YUA	Foursquare	YUAN et al. (2013)
5	TWT_EIN	Twitter	EISENSTEIN et al. (2010)
6	TWT_CHE	Twitter	CHENG et al. (2011)

**Tabela 4.1:** Bases de Dados utilizadas

A Tabela 4.2 mostra a quantidade de *check-ins* por mês para as bases de dados do Brightkite (i.e. BKT\_CHO) e do Gowalla (i.e. GOW\_CHO e GOW\_SCE). Nota-se que a captura dos *check-ins* de BKT\_CHO se iniciou em março de 2008 e foi finalizada em outubro de 2010, totalizando 4.314.673 *check-ins*, realizados por 42.871 usuários em 346.120 locais. GOW\_CHO iniciou a captura de *check-ins* em fevereiro de 2009 e finalizou em outubro de 2010. Neste caso, a quantidade de *check-ins* totaliza 5.938.891 *check-ins* que foram realizados por 98.634 usuários em 782.910 locais. A base GOW\_SCE, por sua vez, iniciou a captura de *check-ins* a partir de janeiro de 2009 e se prolongou até agosto de 2010. Esta base, por sua vez, totaliza uma quantidade de 9.690.477 *check-ins* realizados por 166.686 usuários em 1.083.944 locais.

Para as bases de dados da Tabela 4.2, cada *check-in* é composto por um identificador único de *check-in*, um identificador do usuário que realizou aquele *check-in*, um instante de tempo em que o *check-in* foi realizado e, finalmente, o identificador e as coordenadas geográficas do local em que o *check-in* foi realizado. Além disso, essas bases também disponibilizam os laços sociais dos usuários.

Com o objetivo de eliminar dados irrelevantes e redundantes e, conseqüentemente, otimizar o processamento das análises realizadas e das redes geradas, foram removidos todos os usuários que realizaram menos de cinco *check-ins* e, também, todos os locais em que menos do que cinco *check-ins* foram realizados. Do mesmo modo, observou-se que mais de 60% dos *check-ins* foram capturados no território dos Estados Unidos. Assim, do total de *check-ins* de cada uma das bases da Tabela 4.2, foram considerados 2.643.733 *check-ins* de BKT\_CHO, 3.378.966 de GOW\_CHO e, finalmente, 6.037.727 de GOW\_SCE.

Mês	BKT_CHO			GOW_CHO		GOW_SCE	
	2008	2009	2010	2009	2010	2009	2010
<b>Jan</b>	-	207795	102505	-	294616	8	613243
<b>Fev</b>	-	201315	89037	7	284434	120	616009
<b>Mar</b>	85	225614	90015	897	511445	2149	1043857
<b>Abr</b>	9037	202428	69420	560	513903	2341	1123718
<b>Mai</b>	107855	209657	49132	980	491247	3256	1463512
<b>Jun</b>	118679	200490	94336	660	540341	2940	1709484
<b>Jul</b>	122405	214547	104390	1058	647693	5080	1655108
<b>Ago</b>	120744	209635	118103	1464	784943	5933	736522
<b>Set</b>	111787	188003	154127	7495	882396	15031	-
<b>Out</b>	160603	134053	62250	36693	661140	66237	-
<b>Nov</b>	200940	125104	-	83878	-	171593	-
<b>Dez</b>	191864	118718	-	193041	-	454336	-

**Tabela 4.2:** Quantidade de *check-ins* de BKT\_CHO, GOW\_CHO e GOW\_SCE

Os dados da base do Foursquare (i.e. 4QR\_YUA) foram coletados entre agosto de 2010 e julho de 2011. Seus autores omitiram a data e, no lugar desta, publicaram apenas o instante de tempo do *check-in*. Para cada dia de captura, atribuíram um valor inteiro tornando, assim, impossível ter conhecimento sobre qual data, exatamente, um dado *check-in* foi realizado. A Tabela 4.3 apresenta um exemplo de *check-ins* da referida base.

id usuário	id local	instante de tempo	dia
<b>689</b>	3970	04:30	0
<b>689</b>	3971	08:29	16
<b>2103</b>	1843	12:29	17
<b>2103</b>	1844	08:23	25
<b>2563</b>	1856	08:21	26
<b>2587</b>	1521	09:22	28

**Tabela 4.3:** Exemplos de *check-ins* de 4QR\_YUA

Nota-se, portanto, que da forma como foram publicados os *check-ins* de 4QR\_YUA, não há como se identificar a data completa em que o *check-in* foi publicado, o que impossibilitou alguns tipos de análises, como será descrito adiante. Essa base apresenta uma quantidade de 194.108 *check-ins* sendo distribuídos em 378 dias, realizados por 2.321 usuários que visitaram 5.596 locais. Como os dados foram rotulados apenas com a hora, o minuto e um valor inteiro representando o dia em que o *check-in* foi publicado, a Tabela 4.4 apresenta esses *check-ins* agrupados em somatórios de *check-ins* a cada 30/31 dias de acordo com o referido valor inteiro, i.e. o mês 1 agrupa os *check-ins* publicados nos primeiros 31 dias da base de dados; o mês 2, por sua vez, os *check-ins* publicados nos 30 dias seguintes da base e assim sucessivamente.

De acordo com os autores de 4QR\_YUA, todos os usuários que realizaram *check-in* em menos de cinco locais e todos os locais que tiveram menos de cinco *check-ins* foram removidos, tendo sido disponibilizada, portanto, a base de dados após esse processo de limpeza. Cada

Mês	Check-ins
1	5076
2	7373
3	9508
4	11695
5	11786
6	24565
7	24953
8	22326
9	22148
10	19478
11	19738
12	15462

**Tabela 4.4:** Quantidade de *check-ins* do 4QR\_YUA

*check-in* da base 4QR\_YUA foi disponibilizado contendo as seguintes informações: o dia em que o *check-in* foi realizado (um valor inteiro), o instante de tempo em que ocorreu o *check-in* (no formato *hora:minuto:segundo*), um identificador do usuário que realizou o *check-in*, um identificador do local em que ocorreu o *check-in*, bem como a latitude e a longitude do referido local. Os autores desta base de dados não disponibilizaram os laços sociais dos usuários.

A base de dados TWT\_EIN contém um total de 172.975 tweets. Tecnicamente, tweets não são considerados *check-ins*, já que enquanto que a principal informação de um *check-in* é a visita do usuário a um local, a de um tweet é o texto de 140 caracteres que o usuário publica. Entretanto, tweets geo-localizados são considerados aqui simplesmente como *check-ins*.

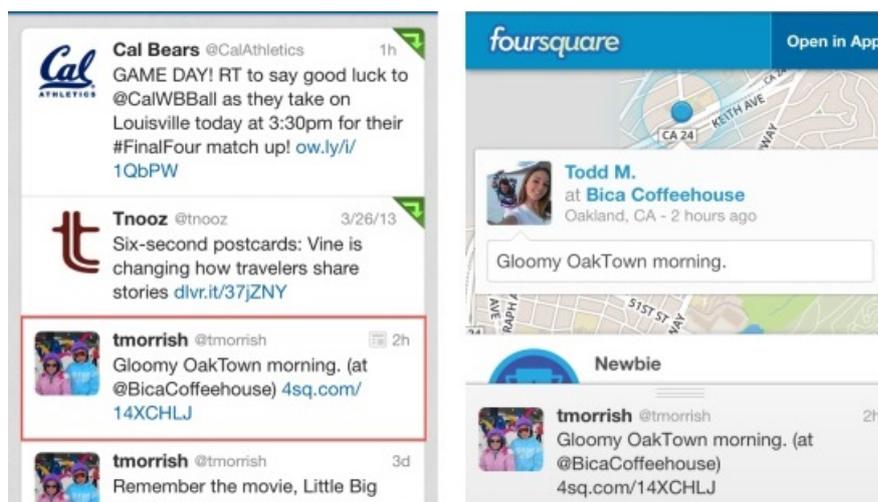
Diante disso, analisando-se a base de dados, pode-se observar que os tweets geo-localizados foram capturados durante a primeira semana de março de 2010. Essa captura se deu, apenas, durante seis dias da semana. A Tabela 4.5 apresenta o quantitativo de tweets por dia. Os autores da referida base de dados publicaram-na de modo a apresentar apenas usuários que compartilharam, pelo menos, 20 mensagens durante o período de coleta, tendo sido esses tweets gerados por pessoas que possuíssem, no máximo, 1000 “seguidores”. Assim, a referida base de dados apresenta um quantitativo de 4.078 usuários que postaram tweets em 35.884 locais.

Dia	Total
02/03/2010	2406
03/03/2010	34516
04/03/2010	38548
05/03/2010	37530
06/03/2010	30093
07/03/2010	29882

**Tabela 4.5:** Quantidade de tweets do TWT\_EIN

Embora cada tweet possa ser rotulado com as coordenadas geográficas do usuário, a API do Twitter não publica outros identificadores dos locais que não sejam as referidas coordenadas.

Em uma rede social baseada em localização isso não ocorre porque cada um dos locais possui um perfil na rede e, conseqüentemente, um identificador único. Essa explicação é necessária porque, frequentemente, dados de localização capturados pelo Twitter foram gerados, na verdade, em alguma rede social do tipo LBSN. Esse é o caso, por exemplo, dos dados da base TWT\_EIN. Isso significa que os usuários apresentados em TWT\_EIN conectaram seus perfis das diferentes redes e, assim, foram rastreados. A Figura 4.2 ilustra um exemplo. É por esta razão, provavelmente, que além de possuírem as coordenadas geográficas dos locais em que foram publicados, os tweets de TWT\_EIN possuem também identificadores dos referidos locais.



**Figura 4.2:** Exemplo de *check-in* publicado originalmente no Foursquare e republicado no Twitter.

A Tabela 4.6 apresenta a quantidade de *check-ins* da base TWT\_CHE por mês e por ano. Pode-se observar, neste caso, que a base de dados iniciou a captura de tweets em fevereiro de 2010 e foi finalizada em fevereiro de 2011, totalizando, assim, 22.390.262 tweets realizados por 224.803 usuários e rotulados com identificadores de 530.374 locais de todo o mundo. Assim como no caso anterior, de acordo com os autores de TWT\_CHE, a referida base é composta por tweets públicos gerados, originalmente, em redes sociais baseadas em localização. Desta forma, além do identificador de usuário que publicou o tweet, o mesmo é composto, também, pelo texto de 140 caracteres publicado pelo usuário, um identificador para o local em que o tweet foi publicado e o conjunto de coordenadas geográficas do referido local.

Dentre os dados de TWT\_CHE, pode-se observar que cerca de 8.462.135 dados de *check-ins* foram rotulados com coordenadas geográficas diferentes, embora tivessem o mesmo identificador de local. Esses *check-ins* foram removidos devido a esta inconsistência. Além disso, todos os usuários que realizaram *check-ins* em menos de cinco locais e todos os locais que tiveram menos de cinco *check-ins*, também, foram removidos. Desta forma, tendo a base TWT\_CHE passado por esse processo de limpeza, o tamanho resultante da mesma passou a ser de 13.093.128 *check-ins* publicados por 156.915 usuários em 137.096 locais.

Por fim, nota-se que a distribuição dos *check-ins* de todas as bases de dados é caracterizada

Mês	2010	2011
<b>Jan</b>	-	1149479
<b>Fev</b>	22660	-
<b>Mar</b>	398534	-
<b>Abr</b>	614262	-
<b>Mai</b>	930148	-
<b>Jun</b>	1318961	-
<b>Jul</b>	1900910	-
<b>Ago</b>	2643670	-
<b>Set</b>	3673706	-
<b>Out</b>	4125829	-
<b>Nov</b>	2931463	-
<b>Dez</b>	2680640	-

**Tabela 4.6:** Quantidade de *check-ins* do TWT\_CHE

por possuir uma quantidade menor de *check-ins* nos primeiros dias ou meses de captura. É importante fazer esta observação para que os resultados apresentados a seguir sejam melhor interpretados a partir do conhecimento de como os *check-ins* são distribuídos temporalmente. Com o objetivo de resumir as informações detalhadas nesta seção, a Tabela 4.7 sumariza as principais informações sobre cada uma das bases de dados utilizadas no decorrer deste trabalho.

Base	Laços Sociais	Usuarios	Locais	Check-ins	Período
BKT_CHO	Presentes	2.643.733	2.643.733	2.643.733	Mar/2008 a Out/2010
GOW_CHO	Presentes	3.378.966	3.378.966	3.378.966	Fev/2008 a Out/2010
GOW_SCE	Presentes	6.037.727	6.037.727	6.037.727	Jan/2008 a Ago/2010
4QR_YUA	Ausentes	193.026	193.026	193.026	-
TWT_EIN	Ausentes	171.487	171.487	171.487	Mar/2010 a Mar/2010
TWT_CHE	Ausentes	6.494.059	6.494.059	6.494.059	Fev/2010 a Jan/2011

**Tabela 4.7:** Resumo das informações presentes em cada base de dados.

## 4.5 Resultados

Esta seção apresenta resultados de uma análise estatística e temporal dos encontros com o objetivo de auxiliar na compreensão de como este evento está distribuído, seja por usuário que participa do encontro, seja por local em que ocorreu o encontro ou em relação ao tempo. Além disso, a descrição da probabilidade de ocorrência de encontro permite que se tenha conhecimento sobre quais leis estatísticas governam a publicação de encontro. A análise apresentada nesta seção possibilitará um melhor entendimento sobre o modelo proposto no Capítulo 5.

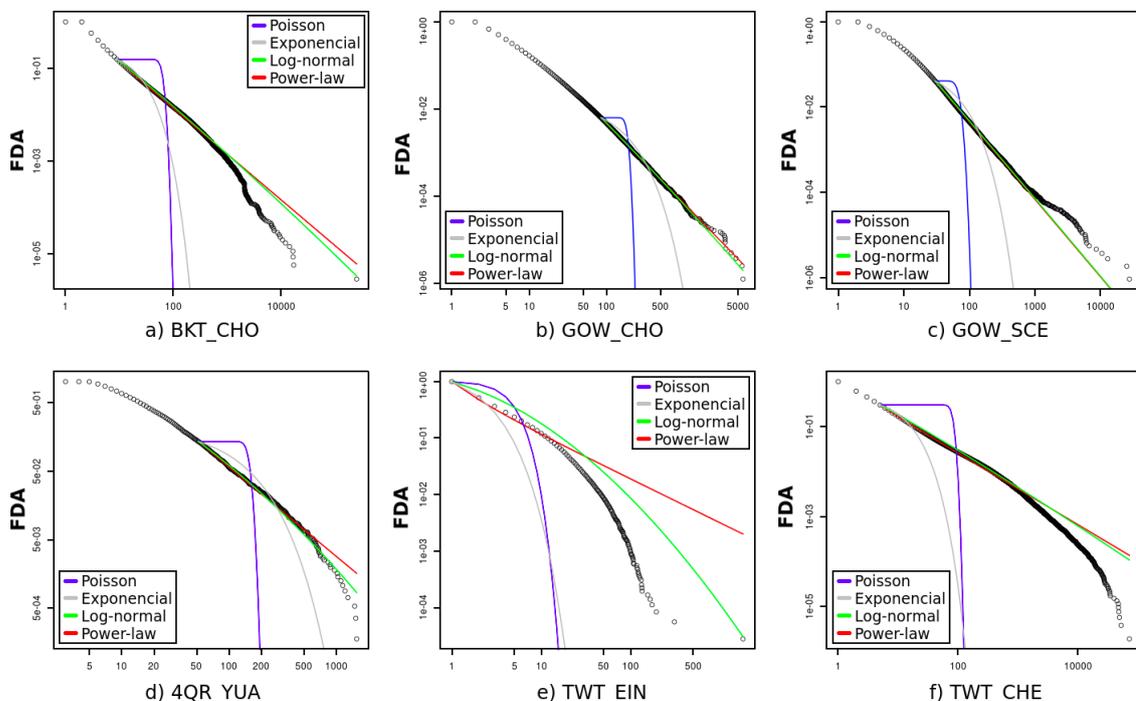
Para uma melhor compreensão do comportamento dos usuários em relação à publicação desses encontros, esta seção inicia com uma descrição da probabilidade dos *check-ins* dos usuários independentemente se o *check-in* faz parte de um encontro ou não (Seção 4.5.1 e Seção 4.5.2, respectivamente). A Seção 4.5.3 exibe uma análise estatística dos encontros, a Seção 4.5.4

descreve como os encontros estão distribuídos em relação ao tempo e, finalmente, a Seção 4.5.5 apresenta uma análise relativa à duração dos encontros.

#### 4.5.1 Análise da Distribuição dos Check-ins

O objetivo da análise apresentada nesta seção é descobrir a distribuição que melhor se ajusta à publicação de *check-ins*. Esta análise foi realizada sob dois pontos de vistas: (1) análise do número de *check-ins* por local onde houve o *check-in* e (2) análise do número de *check-ins* por usuário que publicou o *check-in*.

A Figura 4.3 apresenta a Função de Distribuição Acumulada (FDA) do número de *check-ins* por locais e a aproximação de quatro modelos teóricos (i.e. *power law*, exponencial, log-normal e Poisson). Nota-se que para a maioria dos casos, tanto a curva referente à *power-law* como a curva da log-normal parecem estar bem ajustadas à amostra de dados.



**Figura 4.3:** Comparação entre distribuições teóricas e o número de *check-ins* por locais.

A análise gráfica preliminar sobre os dados de *check-ins* por locais permite notar que os modelos de *power-law* e log-normal são os que mais se aproximam das amostras de dados. Entretanto, como a análise gráfica não é suficiente para que se tome as conclusões devidas, se faz necessário o uso de um teste de inferência estatística para se decidir sobre qual distribuição melhor representa o conjunto de dados estudado. Desta forma, foi utilizado o teste de Vuong, apresentado no Capítulo 3, para comparar a aproximação das amostras de dados às principais funções aqui levantadas, i.e. *power-law*, log-normal e exponencial.

A Tabela 4.8 apresenta os resultados do teste para as amostras de *check-ins* distribuídos por local. Esta tabela exhibe, para cada amostra de dados, o *p-value* da aproximação da referida

amostra com uma função *power-law* ( $p_{pl}$ ) e as métricas definidas pelo teste para auxiliar na comparação entre duas distribuições (neste caso,  $\mathcal{LR}_{pl-ln}$  e  $p_{pl-ln}$  para comparação entre *power-law* e log-normal e  $\mathcal{LR}_{pl-exp}$  e  $p_{pl-exp}$  para comparação entre *power-law* e exponencial).

Base	$P_{pl}$	$\mathcal{LR}_{pl-ln}$	$P_{pl-ln}$	$\mathcal{LR}_{pl-exp}$	$P_{pl-exp}$	Resultado
BKT_CHO	0.00	-2.64	<b>0.00</b>	6.55	<b>0.00</b>	Log-normal
GOW_CHO	<b>0.23</b>	-0.56	0.56	8.71	<b>0.00</b>	Power-law
GOW_SCE	0.00	2.59	<b>0.00</b>	12.21	<b>0.00</b>	Inconclusivo
4QR_YUA	<b>0.21</b>	0.00	<b>0.07</b>	6.13	<b>0.00</b>	Power-law
TWT_EIN	0.00	-27.75	<b>0.00</b>	25.36	<b>0.00</b>	Log-normal
TWT_CHE	0.00	-3.90	<b>0.00</b>	34.80	<b>0.00</b>	Log-normal

**Tabela 4.8:** Teste de Vuong aplicado sobre dados de *check-ins* por local. Valores relevantes estão em negrito para auxiliar na interpretação dos dados.

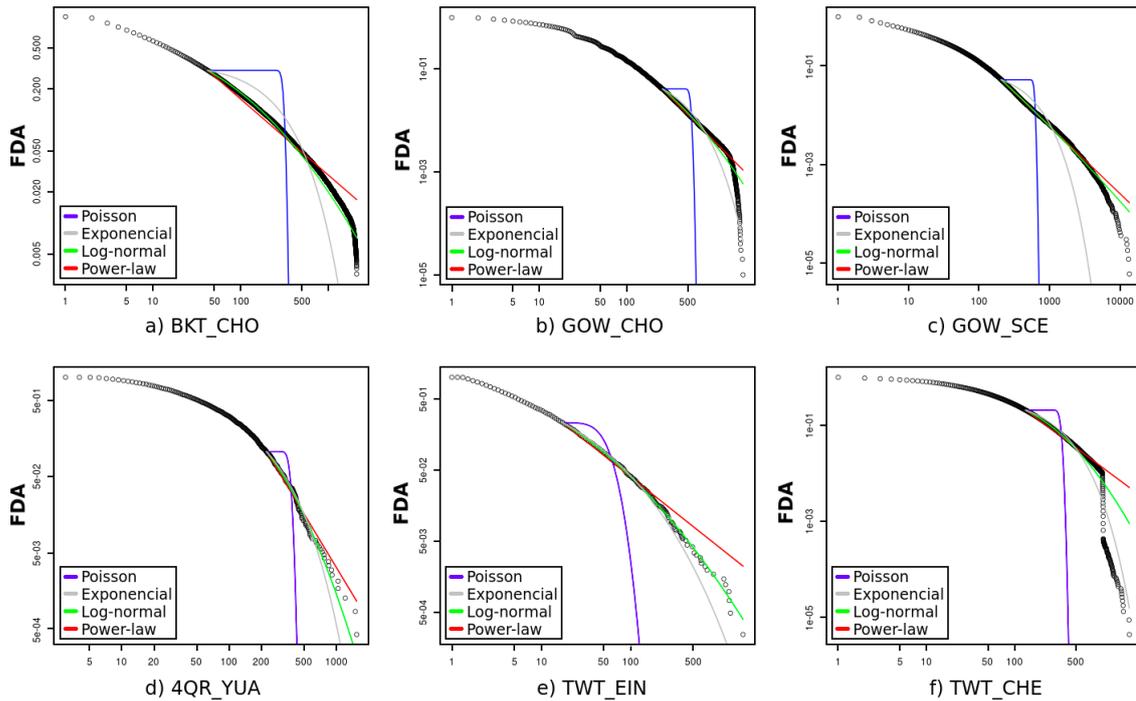
Como o nível de significância foi definido como sendo igual a 90%, se  $p_{pl} > 0.10$ , então, possivelmente, a distribuição de *power-law* se ajusta bem à amostra de dados. Este resultado pode ser confirmado com o Teste de Vuong que, por sua vez, diz que se o valor de  $\mathcal{LR}$  for positivo, então a aproximação da amostra de dados é melhor obtida pela distribuição à esquerda, caso contrário, a distribuição à direita aproxima melhor os dados. Esse resultado pode ser ainda atestado pelo *p-value* do teste de Vuong, i.e. se este *p-value* for menor ou igual ao teste de significância, então o *p-value* corrobora com o valor de  $\mathcal{LR}$ .

Nota-se, portanto, que a função *power-law* parece ser mais adequada para aproximação de GOW\_CHO e 4QR\_YUA. Para as bases BKT\_CHO, TWT\_EIN e TWT\_CHE, a análise mostra que a log-normal se ajusta melhor do que *power-law*. Finalmente, o teste não foi capaz de indicar a função que modela melhor os dados de GOW\_SCE.

A Figura 4.4 apresenta as FDAs do número de *check-ins* por usuário e a aproximação dessa medida com as quatro distribuições já citadas. Pode-se observar que, no caso da distribuição dos *check-ins* por usuários, as distribuições de *power-law*, log-normal e exponencial se ajustam à amostra de dados com mais evidência do que no caso da distribuição dos *check-ins* por local de visitação.

Essa análise gráfica preliminar permite notar que além dos modelos teóricos de *power-law* e log-normal, a exponencial aparenta se ajustar bem a algumas distribuições de *check-ins* por usuários. Este é o caso, por exemplo, dos gráficos da Figura 4.4(d) onde são utilizados os dados provenientes da base 4QR\_YUA, da Figura 4.4(e), na qual foram plotados os dados da base TWT\_EIN e, finalmente, da Figura 4.4(f), onde foram utilizados os dados da TWT\_CHE.

A Tabela 4.9 apresenta o resultado da aplicação do teste de inferência na distribuição de *check-ins* por usuários. Através dessa tabela, pode-se concluir que o modelo log-normal se ajusta bem a maioria dos dados. Para a base TWT\_CHE, na qual é apresentado um (\*), pode-se observar que a comparação entre *power-law* com log-normal indicou que a log-normal seria mais adequada para aproximar a referida amostra de dados, mas ao se comparar *power-law* com exponencial, o mesmo teste mostrou que a função exponencial poderia ser, também,



**Figura 4.4:** Comparação entre distribuições teóricas e o número de *check-ins* por usuários.

uma boa aproximação. Ou seja, o teste mostrou que tanto log-normal como exponencial são melhores aproximações do que a *power-law*. Assim, o mesmo teste foi usado para comparar as aproximações de log-normal e exponencial. Nesta última comparação, pode-se perceber que a função aproximada da log-normal é mais adequada do que a da exponencial.

Base	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	$\mathcal{L}\mathcal{R}_{pl-exp}$	$p_{pl-exp}$	Resultado
BKT_CHO	0.00	-21.90	<b>0.00</b>	24.66	<b>0.00</b>	Log-normal
GOW_CHO	0.00	-6.38	<b>0.00</b>	07.42	<b>0.00</b>	Log-normal
GOW_SCE	0.00	-2.28	<b>0.02</b>	19.92	<b>0.00</b>	Log-normal
4QR_YUA	0.00	-1.62	<b>0.09</b>	0.00	0.99	Log-normal
TWT_EIN	0.00	-3.39	<b>0.00</b>	-1.12	0.26	Log-normal
TWT_CHE	0.00	-40.25	<b>0.00</b>	-8.93	<b>0.00</b>	Log-normal(*)

**Tabela 4.9:** Teste de Vuong aplicado sobre dados de *check-ins* por usuários.

Os resultados da Tabela 4.9 confirmam o que alguns trabalhos da literatura têm reportado: publicações de *check-ins* de usuários podem ser modeladas por meio de uma função log-normal. O diferencial dos resultados descritos aqui é de que além dessa análise ter sido realizada em relação ao número de *check-ins* por usuários, ela ocorreu também em relação ao local em que esses *check-ins* foram publicados (Tabela 4.8). Ou seja, para ambos os casos, a maioria dos dados pode ser modelado por meio de uma função log-normal. Além disso, esta análise ocorreu em uma abrangência de dados maior: foram utilizadas bases de dados que ainda não tinham sido utilizadas neste tipo de estudo (é o caso das bases BKT\_CHO, GOW\_CHO, 4QR\_YUA, TWT\_EIN e TWT\_CHE), reforçando, portanto, o fato de que interações espaço-temporais podem

ser melhor ajustadas por uma função log-normal.

Além de tudo que foi descrito nesta seção, pode-se notar também que a maioria dos check-ins ocorreu em uma minoria de locais e foi publicada, também, por uma minoria de usuários. Pode-se observar, por exemplo, que, no geral, menos de 12% dos locais foram escolhidos por mais da metade dos *check-ins* para todas as bases de dados e que menos de 30% dos usuários realizaram mais do que a metade desses *check-ins*. A Tabela 4.10 apresenta essas relações para as bases BKT\_CHO, GOW\_CHO e GOW\_SCE e a Tabela 4.11, por sua vez, faz o mesmo para as bases 4QR\_YUA, TWT\_EIN e TWT\_CHE.

% de <i>check-ins</i>	BKT_CHO	GOW_CHO	GOW_SCE
+ de 80%	17.64% dos locais	41.84% dos locais	42.37% dos locais
+ de 50%	1.17% dos locais	10.42% dos locais	12.16% dos locais
+ de 80%	14.17% dos usuários	28.41% dos usuários	17.90% dos usuários
+ de 50%	3.93% dos usuários	7.95% dos usuários	3.92% dos usuários

**Tabela 4.10:** Quantidade de *check-ins* postados por usuário e local para BKT\_CHO, GOW\_CHO e GOW\_SCE.

% de <i>check-ins</i>	4QR_YUA	TWT_EIN	TWT_CHE
+ de 80%	34.58% dos locais	30.92% dos locais	3.33% dos locais
+ de 50%	7.90% dos locais	8.42% dos locais	0.35% dos locais
+ de 80%	39.34% dos usuários	64.15% dos usuários	30.36% dos usuários
+ de 50%	15.25% dos usuários	28.69% dos usuários	10% dos usuários

**Tabela 4.11:** Quantidade de *check-ins* postados por usuário e local para 4QR\_YUA, TWT\_EIN e TWT\_CHE.

Embora as relações apresentadas na Tabela 4.10 e na Tabela 4.11 sejam uma indicação de que os dados estão distribuídos de acordo com uma *power-law*, o que foi possível notar através da análise das distribuições de probabilidade apresentadas anteriormente é que esses dados, em sua maioria, são melhor aproximados por meio de uma função log-normal. Essa aparente contradição é apenas semântica, uma vez que tanto a função *power-law* como a função log-normal possuem características equivalentes, conforme demonstrado no Capítulo 3 e que, comumente, um conjunto de dados que poderia ser melhor ajustado por uma log-normal pode ser equivocadamente modelado por uma *power-law*.

Na prática, o fato dos dados de *check-ins* serem melhor ajustados por uma log-normal ao invés de uma *power-law* indica que os usuários demonstram suas relações espaço-temporais (i.e. os *check-ins*) de forma mais regular do que caso esses dados fossem aproximados por uma *power-law*. Além disso, em sendo modelados por uma função log-normal, isso indica que existem menos usuários com alto grau de conexão do que se esses dados fossem modelados por uma *power-law*. Finalmente, conclui-se também que uma quantidade maior de usuários realizam *check-in* de forma mais moderada (ou seja, com um desvio padrão menor do que a média da publicação de *check-ins*) do que se esses dados fossem aproximados por uma *power-law*.

### 4.5.2 Análise Temporal dos Check-ins

Para as bases BKT\_CHO, GOW\_CHO, GOW\_SCE e TWT\_CHE, a informação temporal foi dividida em três grupos diferentes: mês, dia da semana e hora. Para 4QR\_YUA, em que os autores publicaram a base sem informações sobre a data de cada *check-in*, a análise temporal se deu somente pela hora da publicação. Como TWT\_EIN apresenta seus dados distribuídos em uma pequena quantidade de dias (6 dias), essa análise foi feita somente de acordo com a hora de publicação. O principal objetivo da análise temporal é compreender como os dados de *check-ins* são distribuídos de acordo com as referidas unidades temporais e como esta distribuição está relacionada com a análise temporal de encontros, a ser apresentada a seguir, na Seção 4.5.4.

A Figura 4.5 apresenta a distribuição de *check-ins* por mês. Neste caso, estes dados foram normalizados pelo valor máximo do número de *check-ins* para que fosse possível a comparação entre as bases. Pode-se notar que os usuários mais compartilharam os locais que visitaram nos meses de alta temporada (i.e. meses de férias). Os dados do TWT\_CHE apresentam um crescimento do uso da rede entre fevereiro e outubro. Talvez, a explicação para isto se deva a alta heterogeneidade da base de dados no que se refere ao país no qual o tweet foi publicado.

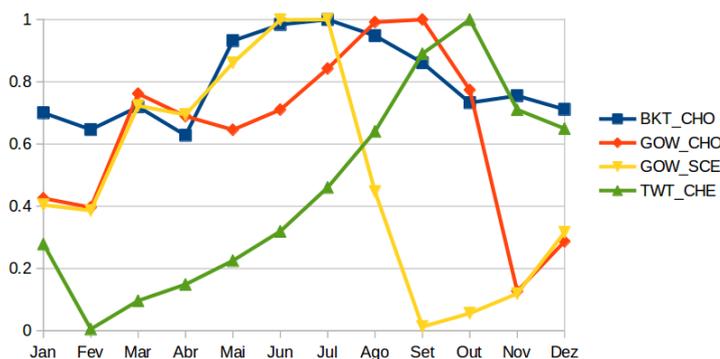


Figura 4.5: Quantidade de *check-ins* por mês.

Da mesma forma, o total de *check-ins* foi distribuído por dias da semana. A Figura 4.6 apresenta os dados normalizados pelo valor máximo para efeito de comparação entre todas as bases. Neste caso, como esperado, a figura mostra a existência de um alto número de *check-ins* em finais de semana e dias próximos a finais de semana (e.g. quintas e sextas).

A Figura 4.7 do lado esquerdo mostra o total de *check-ins* distribuídos por hora do dia para os dados do Brightkite e Gowalla, já a Figura 4.7 do lado direito exibe a mesma medida para os dados publicados no Foursquare e Twitter. Nestes gráficos, pode-se observar que as pessoas compartilham sua localização com uma maior frequência em horários noturnos, particularmente entre 22 horas e 2 horas da madrugada. Observa-se que nos dados provenientes de redes do tipo LBSNs, esta tendência é ainda mais clara, com exceção, apenas, dos dados de 4QR\_YUA.

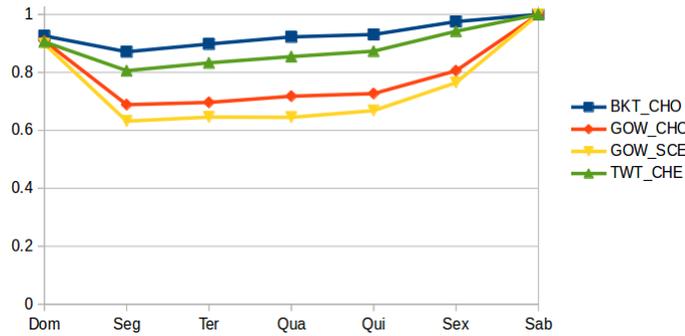


Figura 4.6: Quantidade de *check-ins* por dia da semana.

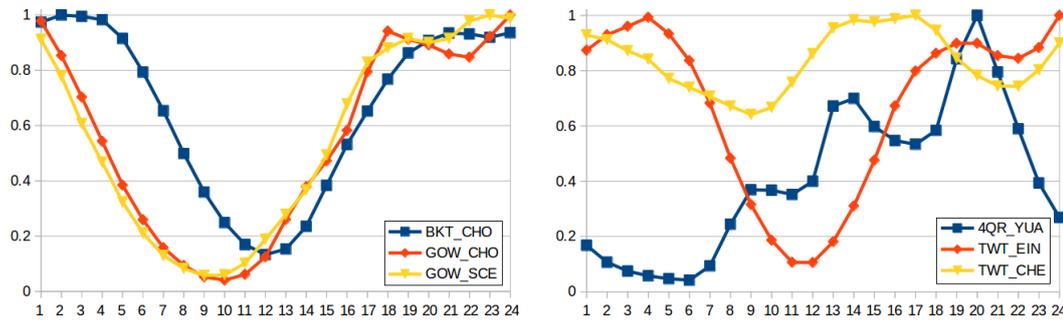


Figura 4.7: Quantidade de *check-ins* por hora do dia.

### 4.5.3 Análise da Distribuição dos Encontros

Para análise da distribuição de encontros, o Algoritmo 1 foi executado passando-se os *check-ins* provenientes das bases de dados descritas na Seção 4.4. Nesta análise, foram passados parâmetros de modo que pudessem ser considerados encontros sociais e encontros não sociais a partir de quatro intervalos de encontro diferentes: 15 minutos, 30 minutos, 1 hora e 2 horas. Essa análise foi feita em quatro amostras diferentes: (1) número de encontros por pares de usuários; (2) número de encontros por usuários *precursores*, ou seja, aqueles usuários que realizaram o primeiro *check-in* de caracterização do encontro; (3) número de encontros por usuários *seguidores*, que são aqueles que realizaram o *check-in* depois dos precursores e, finalmente, (4) número de encontros por locais visitados.

A Tabela 4.12 mostra os resultados do teste de Vuong para o número de encontros sociais distribuídos por pares de usuários. Neste caso, pode-se observar que o modelo *power-law* é o que melhor se ajusta a todas as amostras de encontros sociais com exceção de três. Essas três exceções ocorrem para a base GOW\_SCE, quando  $\Delta t = 30$  minutos,  $\Delta t = 1$  hora e  $\Delta t = 2$  horas.

A Tabela 4.13 apresenta o resultado do teste de Vuong sobre os dados de encontros sociais por usuários precursores e a Tabela 4.14 mostra o mesmo teste para a distribuição de encontros sociais por usuários seguidores. Nota-se que tanto para a distribuição do número de encontros sociais por precursores como para seguidores, a função log-normal é a que melhor se ajusta às amostras de dados. O único caso em que, para ambos, foi apresentada inconsistência no teste, foi para o número de encontros sociais de 1 hora de intervalo de encontro para a base

Base	$\Delta t$	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	Resultado
BKT_CHO	15 min	<b>0.17</b>	-0.61	0.54	Power-law
BKT_CHO	30 min	<b>0.88</b>	-0.40	0.69	Power-law
BKT_CHO	1h	<b>0.48</b>	-0.15	0.88	Power-law
BKT_CHO	2h	<b>0.56</b>	-0.16	0.87	Power-law
GOW_CHO	15 min	<b>0.36</b>	-1.17	0.24	Power-law
GOW_CHO	30 min	<b>0.28</b>	-1.24	0.21	Power-law
GOW_CHO	1h	<b>0.21</b>	-1.47	0.14	Power-law
GOW_CHO	2h	<b>0.31</b>	-1.08	0.28	Power-law
GOW_SCE	15 min	<b>0.23</b>	-0.54	0.58	Power-law
GOW_SCE	30 min	0.00	-3.29	<b>0.00</b>	Log-normal
GOW_SCE	1h	0.00	-8.97	<b>0.00</b>	Log-normal
GOW_SCE	2h	0.00	-8.20	<b>0.00</b>	Log-normal

**Tabela 4.12:** Teste de Vuong para o número de encontros sociais distribuídos por pares de usuários.

GOW\_SCE. Neste caso, o teste indicou que o número de encontros sociais para precursores pode ser modelado por uma função *power-law* e no caso do número de encontros sociais para seguidores o teste se mostrou inconclusivo.

Base	$\Delta t$	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	Resultado
BKT_CHO	15 min	0.00	-2.39	<b>0.01</b>	Log-normal
BKT_CHO	30 min	0.00	-2.96	<b>0.00</b>	Log-normal
BKT_CHO	1h	0.02	-2.50	<b>0.01</b>	Log-normal
BKT_CHO	2h	0.01	-2.53	<b>0.01</b>	Log-normal
GOW_CHO	15 min	0.00	-2.80	<b>0.00</b>	Log-normal
GOW_CHO	30 min	0.00	-3.33	<b>0.00</b>	Log-normal
GOW_CHO	1h	0.00	-3.16	<b>0.00</b>	Log-normal
GOW_CHO	2h	0.00	-2.94	<b>0.00</b>	Log-normal
GOW_SCE	15 min	0.00	-3.15	<b>0.00</b>	Log-normal
GOW_SCE	30 min	0.00	-3.31	<b>0.00</b>	Log-normal
GOW_SCE	1h	<b>0.63</b>	-1.08	0.27	Power-law
GOW_SCE	2h	0.00	-2.25	<b>0.02</b>	Log-normal

**Tabela 4.13:** Teste de Vuong para o número de encontros sociais distribuídos por usuários precursores.

A análise da distribuição de encontros por locais é mostrada na Tabela 4.15 que exhibe o resultado da aplicação do teste de Vuong sobre o número de encontros sociais por locais de ocorrência de encontros. Pode-se observar que, em geral, a distribuição de encontros sociais por local é melhor aproximada por uma *power-law*, com exceção de duas amostras de dados em que o teste se mostrou inconclusivo: encontros sociais ocorridos em 15 minutos para a base BKT\_CHO e para a base GOW\_SCE.

Para as análises apresentadas até aqui, foram considerados apenas os encontros sociais. No que diz respeito aos encontros não sociais, seja em relação à distribuição por pares de

Base	$\Delta t$	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	Resultado
BKT_CHO	15 min	0.00	-3.82	<b>0.00</b>	Log-normal
BKT_CHO	30 min	0.00	-2.89	<b>0.00</b>	Log-normal
BKT_CHO	1h	0.00	-3.03	<b>0.00</b>	Log-normal
BKT_CHO	2h	0.00	-5.41	<b>0.00</b>	Log-normal
GOW_CHO	15 min	0.01	-2.31	<b>0.02</b>	Log-normal
GOW_CHO	30 min	0.00	-3.14	<b>0.00</b>	Log-normal
GOW_CHO	1h	0.00	-3.17	<b>0.00</b>	Log-normal
GOW_CHO	2h	0.00	-3.04	<b>0.00</b>	Log-normal
GOW_SCE	15 min	0.00	-1.82	<b>0.06</b>	Log-normal
GOW_SCE	30 min	0.03	-2.06	<b>0.03</b>	Log-normal
GOW_SCE	1h	<b>0.14</b>	-1.62	<b>0.10</b>	Inconclusivo
GOW_SCE	2h	0.00	-1.62	<b>0.10</b>	Log-normal

**Tabela 4.14:** Teste de Vuong para o número de encontros sociais distribuídos por usuários seguidores.

Base	$\Delta t$	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	Resultado
BKT_CHO	15 min	0.03	-0.60	0.54	Inconclusivo
BKT_CHO	30 min	<b>0.15</b>	-0.87	0.37	Power-law
BKT_CHO	1h	<b>0.86</b>	-0.16	0.86	Power-law
BKT_CHO	2h	<b>0.24</b>	-0.42	0.67	Power-law
GOW_CHO	15 min	<b>0.77</b>	0.06	0.94	Power-law
GOW_CHO	30 min	<b>0.54</b>	-0.49	0.61	Power-law
GOW_CHO	1h	<b>0.87</b>	0.11	0.90	Power-law
GOW_CHO	2h	<b>0.14</b>	0.10	0.91	Power-law
GOW_SCE	15 min	0.00	-1.37	0.16	Inconclusivo
GOW_SCE	30 min	<b>0.27</b>	0.08	0.93	Power-law
GOW_SCE	1h	<b>0.46</b>	-0.30	0.76	Power-law
GOW_SCE	2h	<b>0.37</b>	-0.27	0.78	Power-law

**Tabela 4.15:** Teste de Vuong para o número de encontros sociais distribuídos por locais.

usuários, seja na distribuição deste tipo de encontro por usuários precursores, seguidores ou por locais de ocorrência de encontros, não foi possível concluir sobre qual modelo melhor se ajusta a cada amostra de dados utilizada. Os resultados completos do teste de Vuong para a distribuição de encontros não sociais são listados no Apêndice A.

Em relação às quantidades de encontros publicados, pode-se notar que a maioria dos pares de usuários foi responsável por uma minoria dos encontros não sociais publicados. A Tabela 4.16 mostra esses resultados. Nota-se que para as bases BKT\_CHO, TWT\_EIN e TWT\_CHE, menos de 50% dos pares de usuários publicaram a maioria dos encontros não sociais, enquanto que para as bases GOW\_CHO, GOW\_SCE e 4QR\_YUA, uma quantidade mais moderada de pares de usuários foi responsável pela publicação desse tipo de encontro.

Para as bases BKT\_CHO, GOW\_CHO, GOW\_SCE, TWT\_EIN e TWT\_CHE, nota-se que mais de 80% dos encontros não sociais foram publicados por cerca de 15% dos usuários,

Base	Pares de Usuários		Usuários		Locais	
	Não Social	Social	Não Social	Social	Não Social	Social
BKT_CHO	35.56%	24.53 %	15.07%	18.13%	3.49%	25.87%
GOW_CHO	68.06%	29.62 %	12.31%	29.36%	3.73%	55.96%
GOW_SCE	69.95%	20.65 %	13.54%	23.19%	2.49%	53.74%
4QR_YUA	74.01%	-	40.46%	-	15.36%	-
TWT_EIN	40.27%	-	9.52%	-	0.10%	-
TWT_CHE	41.16%	-	10.25%	-	0.32%	-

**Tabela 4.16:** Relações percentuais de usuários e locais responsáveis pela publicação de mais de 80% encontros.

sejam eles precursores ou seguidores, enquanto que para 4QR\_YUA esse percentual de usuários sobe para cerca de 40%. Finalmente, o percentual de encontros não sociais varia bastante quando é analisada a distribuição por locais escolhidos para publicação. Neste caso, para todas as bases de dados, menos de 16% dos locais foram escolhidos para publicação de encontros não sociais. Para as bases TWT\_EIN e TWT\_CHE essa proporção é ainda menor: menos de 1% dos locais foram escolhidos para publicação de mais de 80% dos encontros não sociais.

A Tabela 4.16 mostra, ainda, os percentuais de encontros sociais em relação aos usuários e aos locais de ocorrência de encontros: mais de 80% dos encontros sociais se concentram em cerca de 25% dos pares de usuários. Esse percentual é semelhante quando os encontros são analisados de forma individual para os usuários. Já em relação à distribuição dos encontros sociais por locais escolhidos para publicação, nota-se que o percentual varia bastante quando são comparados com os percentuais de encontros não sociais por locais.

Tanto a análise da distribuição de encontros por local e por usuário, quanto a análise da distribuição dos percentuais desses encontros ilustram que existem sinais de que as interações espaço-temporais dos usuários (i.e. os encontros) são caracterizadas pelo efeito de Mateus (BARABASI; ALBERT, 1999): a maioria dos usuários publicaram poucos encontros enquanto que poucos usuários publicaram uma grande quantidade de encontros.

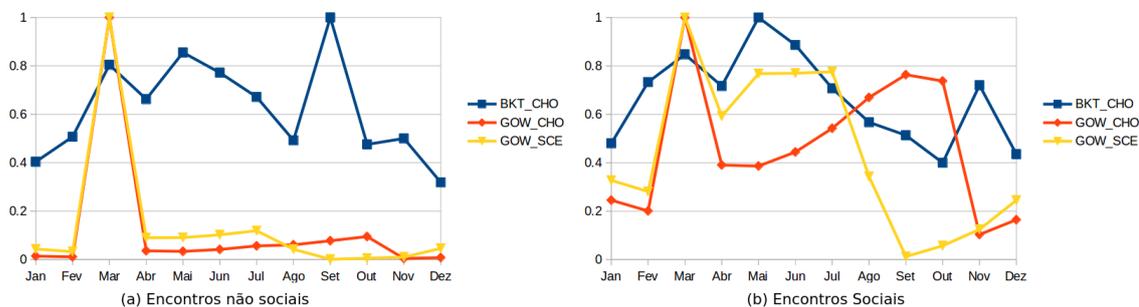
Esta observação é reforçada tanto pelo fato de que, de acordo com estes resultados, encontros podem ser modelados tanto por uma função *power-law* (conforme mostrado na distribuição de encontros por pares de usuário e na distribuição de encontros por local) como por uma função log-normal (nos casos da distribuição de encontros por usuários individuais).

#### 4.5.4 Análise Temporal dos Encontros

A análise temporal dos encontros considerou um intervalo de encontro de até duas horas, uma vez que todos os outros intervalos estudados aqui (i.e. 15 minutos, 30 minutos e 1 hora) acabam por serem considerados também. O objetivo desta análise é compreender como as publicações de encontros estão distribuídas temporalmente e de que forma essa distribuição está relacionada com a distribuição dos *check-ins*.

Os encontros detectados nas bases BKT\_CHO, GOW\_CHO e GOW\_SCE foram agrega-

dos em três grupos diferentes: mês, dia da semana e hora. A Figura 4.8 mostra a distribuição do total de encontros por mês para essas bases. Os dados desta figura foram normalizados pelo valor máximo para que fosse possível a comparação entre diferentes bases de dados.



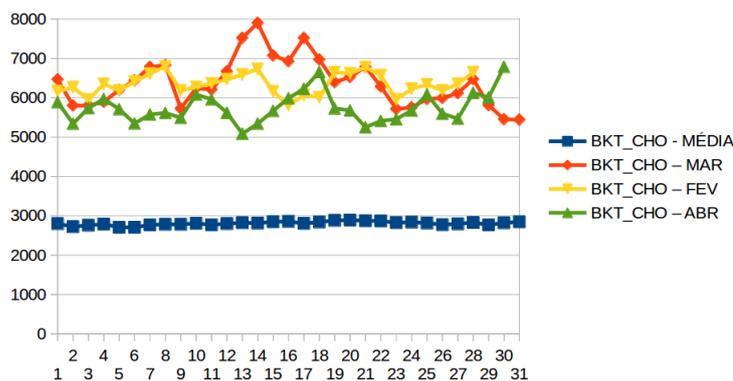
**Figura 4.8:** Quantidade de encontros por mês.

Apesar da base de dados TWT\_CHE possuir *check-ins* distribuídos ao longo dos doze meses, só foram detectados encontros entre os meses de setembro e novembro de 2010. Observando-se a Figura 4.8, nota-se que os encontros não sociais não são caracterizados por possuírem o mesmo padrão temporal dos *check-ins* (i.e. Figura 4.5). Além disso, pode-se perceber que na distribuição temporal dos encontros, sejam eles sociais ou não, ocorreu um grande número de encontros no mês de março para todas as bases estudadas, o que não foi prontamente observado na análise temporal de *check-ins*.

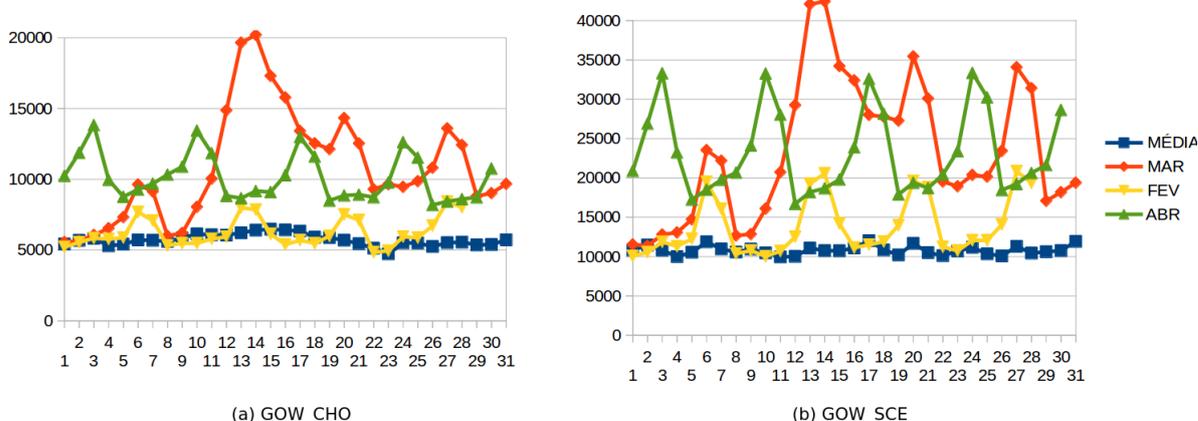
A investigação sobre as possíveis razões do alto número de encontros no mês de março se deu, inicialmente, observando-se que o número de *check-ins* entre a terceira e quarta semana daquele mês aumentou em até 80% em todas as bases de dados. Tanto a Figura 4.9 quanto a Figura 4.10 ilustram o número médio de *check-ins* por dia (curva em azul) e compara esse valor médio com o número de *check-ins* do mês anterior e posterior ao mês em análise, i.e. fevereiro (curva em amarelo) e abril (curva em verde). A Figura 4.9 mostra que, apesar de os meses de fevereiro, março e abril de BKT\_CHO possuírem *check-ins* acima da média, o mês de março, especificamente e, especialmente entre a terceira e quarta semana, apresenta um número que é ainda superior aos demais. A curva referente aos *check-ins* de março é ilustrada em vermelho.

A Figura 4.10(a) mostra o aumento no número de *check-ins* durante a terceira e quarta semana do mês de março para GOW\_CHO e a Figura 4.10(b) ilustra a mesma relação para GOW\_SCE. Pode-se notar que, para essas bases, o aumento é significativamente maior durante as últimas semanas do mês de março quando comparado com o mesmo gráfico da base BKT\_CHO (Figura 4.9). Embora o aumento no número de *check-ins* neste período não tenha se mostrado relevante na análise temporal dos *check-ins* (apresentada na Seção 4.5.2), a análise de encontros permitiu que esse aumento fosse detectado. Como tal aumento ocorreu em todas as bases de dados, foi feita uma investigação mais detalhada a respeito das razões para o alto número de publicações com coincidência espaço-temporal dos usuários.

Assim, percebeu-se que uma grande quantidade destes *check-ins* ocorrera, especificamente, na região do centro de convenções de Austin, TX, nos Estados Unidos, respeitando-se um



**Figura 4.9:** Número médio de *check-ins* por dia comparado a número de *check-ins* para os meses de fevereiro, março e abril para a base BKT\_CHO.



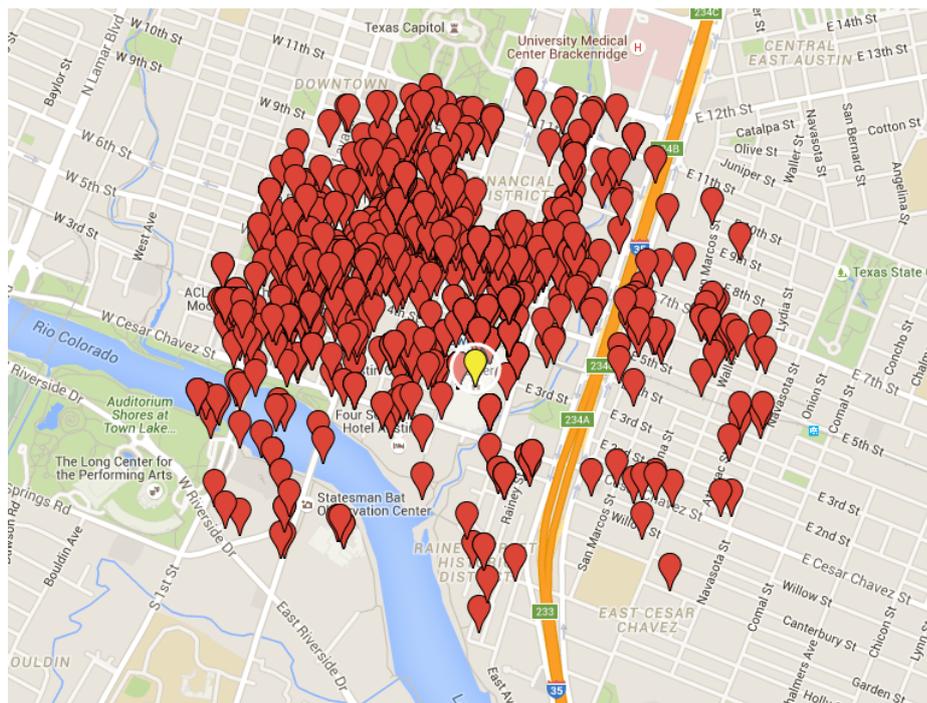
**Figura 4.10:** Número médio de *check-ins* por dia comparado a número de *check-ins* para os meses de fevereiro, março e abril para as bases GOW\_CHO e GOW\_SCE.

raio de abrangência de 1Km a partir do referido centro de convenções. A Figura 4.11 ilustra o centro de convenções de Austin ao centro do mapa com um marcador amarelo e todos os locais da base GOW\_SCE dentro desse raio de abrangência e que foram utilizados para publicação de *check-ins* no período de tempo já citado. Estes locais estão representados por marcadores em vermelho. A imagem foi obtida a partir do Google Maps<sup>1</sup>.

A região em estudo apresentou um aumento entre 800% e 2500% de *check-ins* no mês de março. Esse aumento é exibido na Figura 4.12. Para a base BKT\_CHO, por exemplo, não mais do que 0.5% dos *check-ins* de cada mês ocorrem na região em estudo. Entretanto, para o mês de março, cerca de 3% dos *check-ins* de toda a base de dados ocorreram na referida região. Ou seja, enquanto que, por mês, a região agrega uma média de, aproximadamente, 3500 *check-ins*, no mês de março esse valor subiu para cerca de 43 mil *check-ins* realizados na mesma região.

Em relação as bases do Gowalla, pode-se observar que esse aumento foi ainda mais relevante. Para GOW\_SCE, por exemplo, enquanto que a região em estudo agrega entre 0.7% e 3.5% dos *check-ins* nos outros meses, no mês de março, cerca de 11% de todos os *check-ins* da

<sup>1</sup>maps.google.com.br



**Figura 4.11:** Locais em torno do centro de convenções de Austin utilizados para realização de encontros. Imagem obtida a partir do Google Maps.

base de dados foram publicados naquela região. Esse aumento percentual significa que enquanto que no mês de março foi constatado cerca de 81 mil *check-ins*, nos outros meses esse número produziu uma média de 4.7 mil *check-ins*.

Pode-se concluir que o alto número de encontros detectados na região se deu pelo fato de, tradicionalmente, o centro de convenções ser o local escolhido para sediar o evento “South by Southwest” (SXSW) (MUSICA, 2010) que ocorre, inclusive, sempre entre a segunda e quarta semana de março de cada ano. O SXSW é um conhecido festival de música e cinema que reúne milhares de pessoas e que ocorre há, pelo menos, 30 anos no mesmo local<sup>2</sup>.

Além disso, pode-se perceber que, semelhantemente à distribuição de *check-ins*, o número de encontros sociais ocorre com maior incidência nos meses de alta temporada, i.e. meses de férias, como junho/julho e dezembro/janeiro. Nota-se, portanto, que a dimensão social atua como um importante fator na publicação de encontros, uma vez que a distribuição temporal dos *check-ins* em muito se assemelha à distribuição temporal de encontros sociais.

Para 4QR\_YUA e TWT\_EIN, a análise temporal foi feita somente pela hora da publicação. Com relação à distribuição de encontros por dia da semana, pode-se observar que para ambos tipos de encontros, sejam sociais ou não, a tendência é que os usuários publiquem esses encontros durante finais de semana, como mostra a Figura 4.13. Entretanto, pode-se observar que esta tendência é ainda mais visível quando são considerados os encontros sociais.

Finalmente, embora o número de *check-ins* distribuídos por hora tenha seguido um certo

<sup>2</sup><http://www.sxsw.com/>

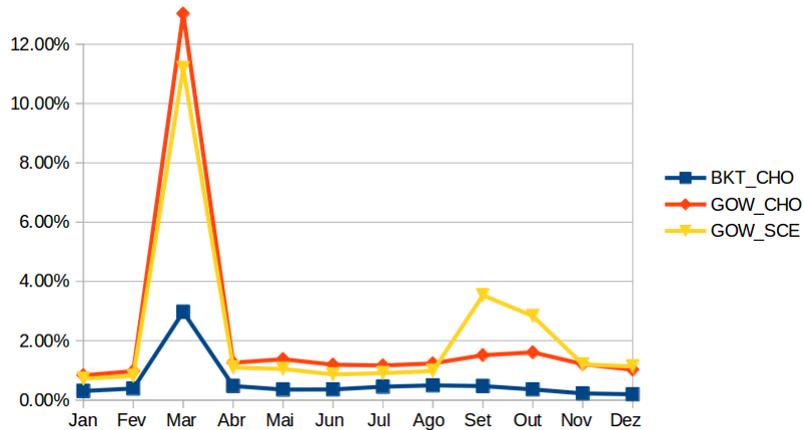


Figura 4.12: Percentual de *check-ins* no Centro de Convenções de Austin.

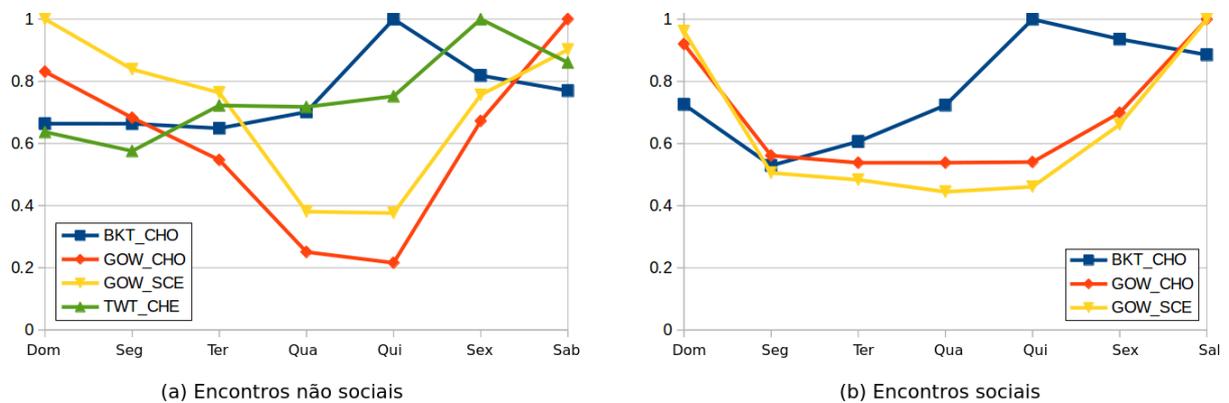


Figura 4.13: Encontros por dia de semana.

padrão de publicação (com um alto número de *check-ins* à noite, por exemplo), foi observado que, embora essa tendência ainda exista, a mesma não é tão evidente quando são analisados os *check-ins* das amostras. A Figura 4.14 apresenta a distribuição de encontros não sociais ao longo do dia enquanto que a Figura 4.15 exibe a distribuição de encontros sociais durante um dia. Nota-se, portanto, que os encontros não sociais têm uma distribuição mais uniforme ao longo do dia enquanto que os encontros sociais apresentam, de fato, a tendência de usuários publicarem seus encontros entre as 23h e 3 horas do dia seguinte.

#### 4.5.5 Duração dos Encontros

A duração de encontros diz respeito à análise do intervalo de tempo entre cada par de *check-ins* dos usuários que fazem parte do encontro. O objetivo aqui é analisar com qual frequência os usuários decidem publicar os seus encontros. No geral, pode-se notar que os encontros não sociais são publicados sem que haja um padrão temporal relativo à duração dos mesmos. Ou seja, independentemente do intervalo de tempo adotado, sempre ocorrerá encontros não sociais entre dois usuários e esse número de encontros varia entre uma base de dados e outra.

Ao contrário dos encontros não sociais, os encontros sociais são caracterizados por

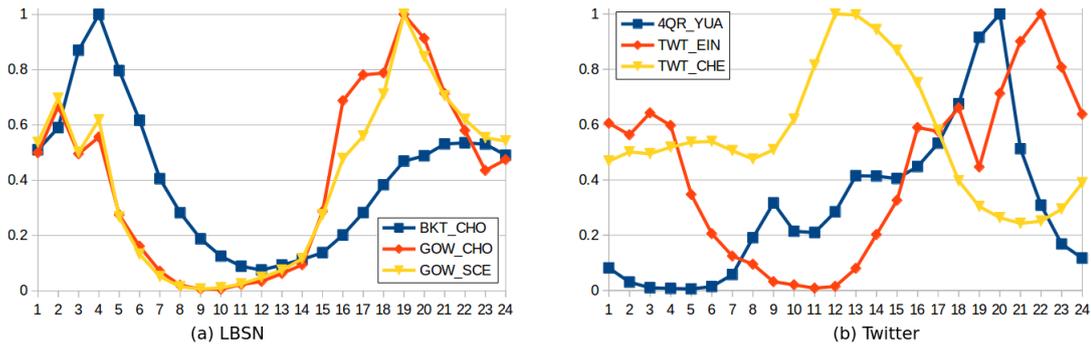


Figura 4.14: Encontros não sociais por hora do dia.

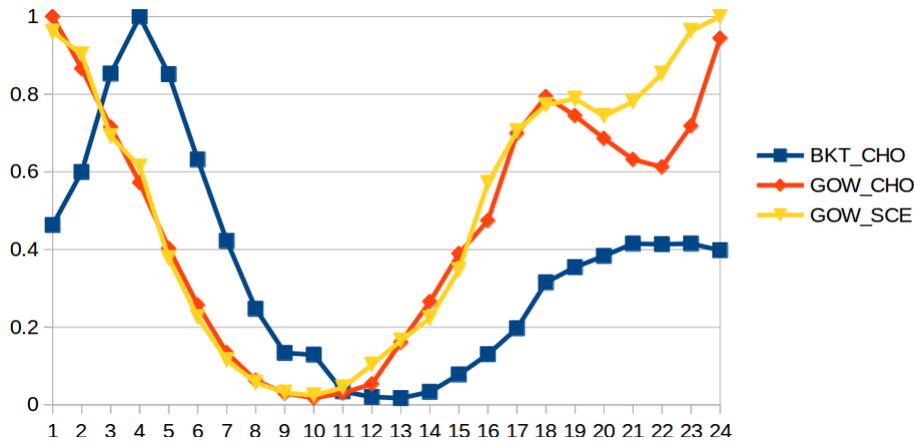


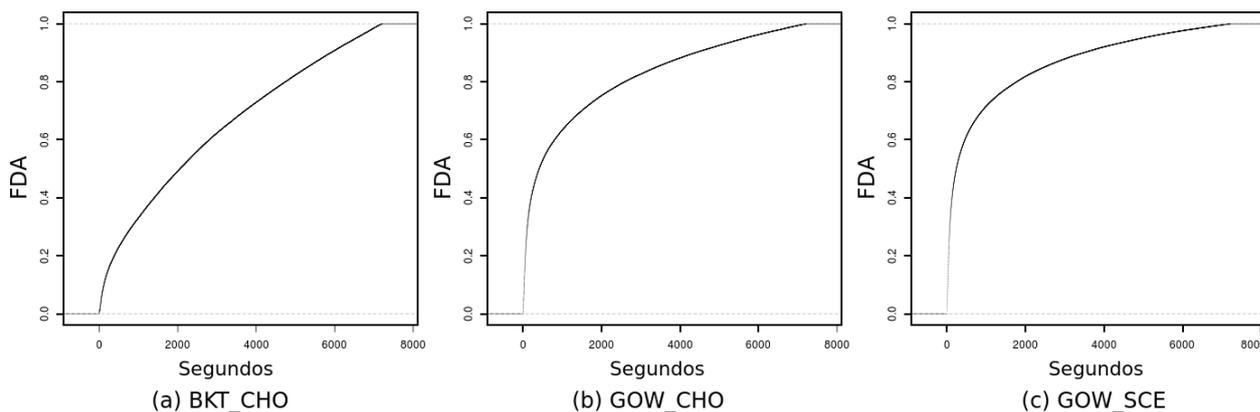
Figura 4.15: Encontros sociais por hora do dia.

serem publicados, frequentemente, em um espaço de encontro delimitado entre os *check-ins* dos usuários envolvidos na publicação dos mesmos e, quanto menor for esse intervalo de encontro, maior a quantidade de encontros sociais publicados. Por exemplo, através da distribuição da função acumulada dos encontros em relação ao intervalo de encontro, pode-se notar para todas as bases de dados com laços sociais (i.e. BKT\_CHO, GOW\_CHO e GOW\_SCE) que cerca de 70% dos encontros sociais foram publicados com um intervalo de encontro menor ou igual a uma hora. A Figura 4.16 apresenta a função acumulada dessas bases de dados.

Analisando-se as FDAs da Figura 4.16, pode-se perceber, portanto, que quanto maior o intervalo de tempo, maior a probabilidade de se ter um encontro social publicado. Isso significa que se um usuário  $A$  publicar um *check-in* em um local  $P$  e, estando com um amigo  $B$  no mesmo local, a probabilidade de  $B$  publicar que está com  $A$  em  $P$  diminui com o passar do tempo.

## 4.6 Considerações Finais

Este capítulo apresentou a caracterização de encontros detectados em seis diferentes bases de dados reais, com o objetivo de compreender as interações espaço-temporais, bem como suas relações com a dimensão social. Os resultados foram descritos quantitativamente e podem ser resumidos como segue.



**Figura 4.16:** FDA de encontros sociais por hora do dia.

- Dados reais de *check-ins* foram comparados com distribuições teóricas e, diferentemente de outros tipos de dados de redes sociais, a conclusão é que dados espaço-temporais são melhor aproximados por uma função log-normal. Esse resultado indica que apesar de possuírem características semelhantes, já que ambas distribuições log-normal e power-law são equivalentes, os dados de *check-ins* são distribuídos de forma mais regular do que os dados meramente sociais;
- Podem-se notar alguns padrões temporais na distribuição dos *check-ins*: a publicação de *check-ins* ocorre mais comumente em meses de férias, como janeiro, junho, julho e dezembro. Do mesmo modo, existem mais publicações de *check-ins* em finais de semana e dias adjacentes a finais de semana, como quintas e sextas do que nos outros dias. Finalmente, pode-se observar que em horários noturnos existe uma maior incidência de publicação de *check-ins*, particularmente entre as 22 horas e 2 horas da madrugada. Em redes consideradas LBSN, esta tendência é ainda mais evidente;
- A análise de distribuição de encontros permitiu concluir que o número de encontros por pares de usuários pode ser bem ajustado, em sua maioria, utilizando-se um modelo baseado em uma *power-law*. Do mesmo modo, a publicação de encontros em locais em que ocorreram esses encontros também pode ser modelada por uma *power-law*. Entretanto, foi observado que uma função log-normal pode ser bem ajustada para a modelagem da publicação individual de encontros por usuários;
- Pode-se perceber o papel dos laços sociais na publicação de encontros quando analisada a relação de encontros por usuários e por locais: enquanto que não foi possível observar uma relação comum para todas as bases entre a quantidade de encontros não sociais publicados por pares de usuários e por locais, pode-se perceber que a maioria dos encontros sociais são publicados por uma minoria de pares de usuários em todas as bases analisadas;

- Pode-se perceber que os encontros não sociais não apresentam o mesmo padrão temporal dos *check-ins*. Ao contrário, tanto a análise de encontros sociais por mês como a análise desses encontros por dia da semana e por hora mostrou que a distribuição temporal desse tipo de encontro em muito se assemelha com a distribuição temporal dos *check-ins*;
- Por meio da análise temporal dos encontros foi possível caracterizar um aglomerado de pessoas que, independentemente da base de dados utilizada, produziram um padrão temporal de publicação no mesmo local e durante o mesmo intervalo de tempo;
- Como esperado, a maioria dos encontros sociais são caracterizados por serem publicados em um intervalo de tempo pequeno entre *check-ins* envolvendo o encontro. Ou seja, os resultados da Seção 4.5.5 indicaram que a intenção de um usuário em publicar um encontro com outro usuário é inversamente proporcional ao passar do tempo entre um *check-in* publicado pelo primeiro usuário e o instante de tempo em que o segundo usuário decide, ou não, publicar o *check-in*;

Em linhas gerais, esse capítulo ilustrou, por meio de um conjunto de dados capturados a partir de diferentes redes sociais, que as interações espaço-temporais entre usuários seguem padrões de distribuição semelhantes. Do mesmo modo, foi detectado em todas as bases analisadas que a publicação dessas interações são impactadas pelo aspecto social entre os usuários. Além disso, pode-se perceber que os encontros seguem um mesmo padrão temporal em todas as bases de dados.

As contribuições desse capítulo podem ser aplicadas em diversos contextos: o conhecimento sobre as leis que governam as interações espaço-temporais podem ser úteis para modelagem de mobilidade de usuários e suas diversas aplicações, na análise de usuários e locais que podem influenciar outros usuários a interagirem espaço-temporalmente, em estratégias de marketing e publicidade e, também, na construção de sistemas de recomendação de locais de encontro ou sistemas de recomendação de encontros, por exemplo.

## 5

### O Modelo SiST

Diante do novo cenário das redes sociais dotadas de serviços de localização, modelos de dados têm sido propostos para representar as diferentes relações entre os usuários, conforme já observado no Capítulo 3. Entretanto, pode-se observar que ainda é um desafio para a área de análise de redes sociais a representação, o processamento e a análise das interações espaço-temporais capturadas pelos usuários deste tipo de rede.

Diante disso, este capítulo propõe o modelo SiST (MORENO; TIMES; MATWIN, 2015), que é baseado na teoria dos grafos, onde os nós representam usuários e cada ligação entre dois nós (i.e. entre dois usuários) representa interações espaço-temporais entre os mesmos. O modelo SiST serve como um novo ambiente de apoio à análise de redes sociais equipadas com informações de localização dos usuários. A partir deste modelo pode-se fazer uso de algoritmos tradicionais da teoria de grafos para se detectar padrões estruturais, fazer uso de algoritmos e técnicas de mineração de dados para realizar previsão de dados, extração de padrões de comportamento de usuários, de mobilidade de usuários, dentre outros.

Propor um modelo como o SiST exige desafios distintos: (1) o modelo deve ser o mais genérico possível para acomodar as diferentes informações provenientes dos diversos tipos de redes sociais bem como as recorrentes mudanças desta área, (2) as redes instanciadas a partir do modelo devem ter baixo custo de memória, (3) os algoritmos de criação destas redes devem ser de baixo nível de complexidade e, por fim, (4) para extração de conhecimento, algoritmos e técnicas de mineração de dados devem poder ser aplicados às redes geradas por este modelo. Para tanto, o modelo é apresentado em diferentes níveis de abstração: modelo conceitual, lógico e físico. O objetivo final da proposta do modelo é auxiliar na busca por respostas às perguntas apresentadas na Seção 1.2.

O modelo SiST é descrito neste capítulo, que está organizado como segue. Na Seção 5.1, é apresentada uma introdução ao capítulo, na qual um cenário hipotético é descrito com o objetivo de auxiliar na compreensão dos conceitos expostos ao longo do mesmo. Na Seção 5.2, o modelo conceitual é proposto, onde a definição formal de uma rede SiST é dada. A Seção 5.3 descreve o modelo lógico de dados, listando seus principais elementos. Na Seção 5.4, o modelo físico de dados é descrito, o qual é apresentado de acordo com uma nova definição de matriz de adjacência e, finalmente, na Seção 5.5 o algoritmo de criação de redes a partir do modelo SiST é

apresentado. As redes SiST são validadas na Seção 5.6 por meio de experimentos que usaram as bases de dados apresentadas no Capítulo 4. Por fim, a Seção 5.7 descreve as considerações finais relativas ao presente capítulo.

## 5.1 Introdução

Esta seção apresenta um cenário hipotético que será utilizado ao longo do capítulo (Seção 5.1.1) bem como introduz possíveis representações de um encontro espaço-temporal entre dois usuários de uma rede social online (Seção 5.1.2).

### 5.1.1 Um Cenário Hipotético

A Tabela 5.1 ilustra um conjunto de *check-ins* hipoteticamente publicados por cinco usuários (identificados na coluna *uid*) e ocorridos em três locais diferentes: POI1, POI2 e POI3. Estes locais são identificados na coluna *pid*. Cada um desses *check-ins* são realizados em um instante de tempo (coluna *timestamp*) e, finalmente, o usuário que publicou o *check-in* pode, ainda, publicar um texto acompanhado a este *check-in*.

<i>i</i>	<i>uid</i>	<i>pid</i>	<i>timestamp</i>	<i>texto</i>
1	1	POI1	10/12/2014 09:35	“Praia com pessoas especiais.”
2	2	POI1	10/12/2014 09:50	“Praia com amigos.”
3	2	POI1	18/12/2014 20:00	“”
4	5	POI1	18/12/2014 20:29	“Confraternização de fim de ano!”
5	1	POI2	20/12/2014 20:15	“Comemorando o sucesso de 2014!”
6	2	POI2	20/12/2014 20:20	“Férias!!!”
7	2	POI3	04/01/2015 19:20	“Reunião de negócios rsrs.”
8	3	POI3	04/01/2015 19:48	“”
9	3	POI2	10/01/2015 20:20	“Começando o ano bem <3.”
10	5	POI2	10/01/2015 20:25	“Com ela.”
11	4	POI1	10/01/2015 11:35	“Primeira praia do ano.”
12	5	POI1	10/01/2015 11:40	“”
13	2	POI1	10/01/2015 12:00	“”
14	5	POI2	15/01/2015 21:00	“Jantar merecido. Ninguém é de ferro.”
15	3	POI2	15/01/2015 21:11	“Jantar ostentação!”
16	4	POI3	21/01/2015 12:15	“Almoço com os de sempre.”
17	3	POI3	21/01/2015 12:18	“”
18	1	POI3	21/01/2015 12:22	“”
19	1	POI2	30/01/2015 20:21	“”
20	4	POI2	30/01/2015 20:40	“”
21	2	POI1	10/01/2015 10:21	“Com ela, no mesmo lugar, desde 2014.”
22	1	POI1	10/01/2015 10:40	“Um mês de namoro onde tudo começou.”

**Tabela 5.1:** Conjunto hipotético de dados de *check-ins*

O cenário da Tabela 5.1 é utilizado ao longo desse capítulo com o objetivo de auxiliar no

entendimento dos conceitos definidos. Pode-se observar, por exemplo, que o usuário de  $uid = 1$ , ou simplesmente usuário 1, publicou cinco *check-ins*, sendo dois deles no local POI1 (i.e. os *check-ins* com  $i = 1$  e  $i = 22$ ), dois em POI2 ( $i = 5$ ,  $i = 19$ ) e, finalmente, um em POI3 ( $i = 18$ ).

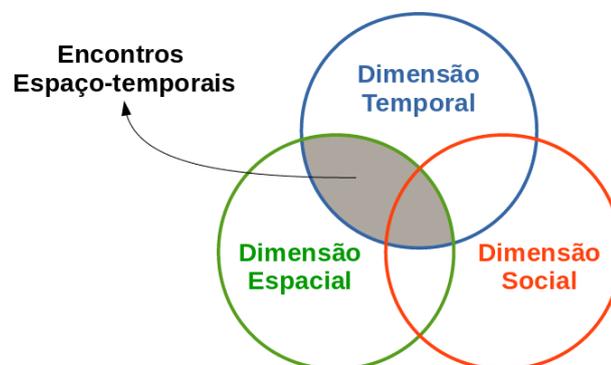
Neste cenário, a rede social em questão é uma rede de amizade bidirecional, ou seja, os usuários são mutuamente considerados amigos. Os laços sociais do referido cenário são apresentados na Tabela 5.2. Por exemplo, o usuário 1 é amigo dos usuários 2, 3 e 4. O usuário de 2, por sua vez, é amigo dos usuários 1, 3 e 4 e assim sucessivamente.

$u1$	$u2$
1	2
1	3
1	4
2	3
2	4
3	5
4	5

**Tabela 5.2:** Conjunto hipotético de dados de amizade

### 5.1.2 Representações de um Encontro Espaço-temporal

A Figura 5.1 ilustra uma visão conceitual das diferentes dimensões de informações presentes em encontros extraídos a partir de redes sociais equipadas com funcionalidade de localização de usuários. Encontros, como definidos no Capítulo 4, são caracterizados por possuírem informações espaciais e temporais. Além disso, caso o encontro tenha ocorrido entre usuários que possuem laços sociais, o encontro também pode ser identificado como um encontro espaço-temporal e social, i.e. um encontro social.

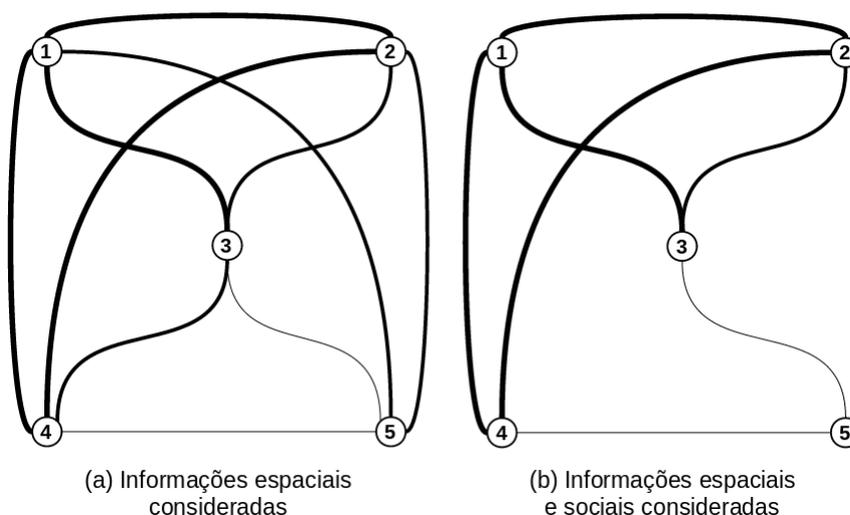


**Figura 5.1:** Visão conceitual das diferentes dimensões de informações presentes em dados de redes sociais.

Diante da revisão da literatura do Capítulo 3, é importante observar que a dimensão social está inerentemente presente em um encontro espaço-temporal mesmo que ambos usuários

não tenham laços sociais expressos na rede social, uma vez que, estando dois usuários no mesmo local, existe uma grande possibilidade desses usuários se conectarem socialmente.

Considerando o cenário exposto na Seção 5.1.1, podem-se criar redes de acordo com diferentes pontos de vistas. Por exemplo, a Figura 5.2(a) ilustra um exemplo de uma rede construída considerando-se apenas as coincidências espaciais entre usuários. Neste caso, como diferentes usuários estiveram no mesmo local, independentemente das circunstâncias temporais ou sociais, estes usuários são conectados. Essa rede hipotética teria um peso nas conexões que estaria relacionado com o número de locais em comum que os usuários visitaram.



**Figura 5.2:** Redes construídas considerando-se (a) informações espaciais e (b) informações espaciais e sociais.

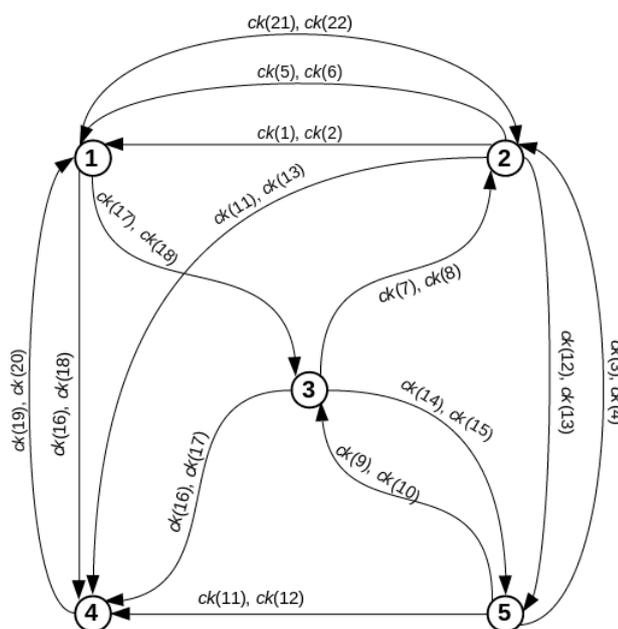
Pode-se notar, portanto, que a conexão entre os nós que representam os usuários 1 e 2 possuem um peso maior do que aquela que conecta os usuários 3 e 5, por exemplo. Isso ocorre porque os usuários 1 e 2 possuem mais coincidências espaciais do que os usuários 3 e 5.

A Figura 5.2(b) ilustra uma rede que faz uso das informações espaciais e sociais do cenário exposto. Neste caso, além de se considerar as coincidências espaciais, dois nós só são conectados se também possuírem laços sociais. Pode-se observar, portanto, que a rede (ou o grafo) da Figura 5.2(b) mostra um número menor de conexões do que a rede à esquerda.

No mesmo sentido, pode-se fazer uso das informações temporais para analisar as coincidências espaço-temporais, ou seja, investigar as situações em que dois usuários estiveram no mesmo local e em um instante de tempo aproximado. Na prática, uma rede que considera esse tipo de evento é, na verdade, uma rede de encontros. Para que essa rede seja construída, deve-se estabelecer um intervalo de tempo que represente o limite máximo entre os instantes de tempo dos *check-ins* envolvidos nos encontros, i.e. um *intervalo de encontro*.

Assim, considerando-se um intervalo de encontro de 30 minutos, pode-se construir uma rede tal qual a ilustrada na Figura 5.3. Esta rede apresenta conexões direcionadas entre cada dois nós. A direção de cada conexão respeita a ordem temporal dos *check-ins* envolvidos no encontro. Por exemplo, entre os usuários 1 e 2 existem três conexões: a primeira, rotulada com “*ck(1)*,”

$ck(2)$ ”, existe porque os referidos usuários realizaram o *check-in* 1 (i.e.  $i = 1$ ) e o *check-in* 2 no mesmo local em um intervalo de tempo de, exatamente, 15 minutos. Como o usuário 2 não foi o primeiro a realizar o *check-in*, a referida conexão tem origem no nó que representa o mesmo e tem destino para o nó que representa o usuário 1; a segunda conexão, rotulada com “ $ck(5), ck(6)$ ”, parte do usuário 2 para o usuário 1 pela mesma razão da conexão anteriormente citada; finalmente, a conexão “ $ck(21), ck(22)$ ” tem o destino contrário das anteriores devido à ordem temporal em que os *check-ins* envolvidos está disposta: primeiramente ocorre o checkin 21, do usuário 2, em *POII* e, 19 minutos depois, o usuário 1 realiza o *check-in* 22 no mesmo local.

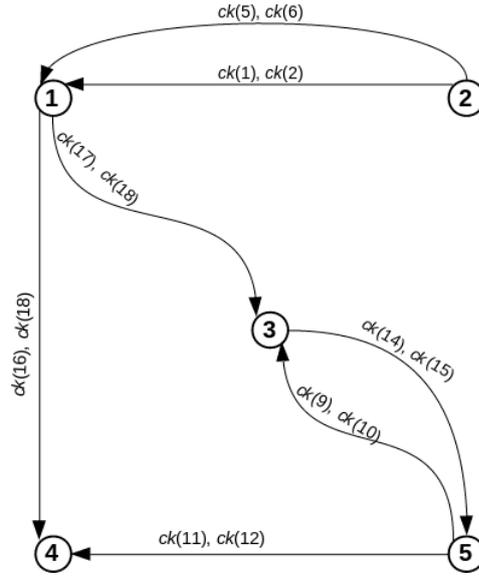


**Figura 5.3:** Redes construídas considerando-se informações espaço-temporais para  $\Delta t = 30$  min.

Finalmente, uma rede de encontros poderia, ainda, ser construída a partir de informações sociais e espaço-temporais. Assim, uma conexão entre dois nós de uma rede estaria representando um encontro ocorrido entre duas pessoas que são consideradas amigas na rede social online. A Figura 5.4 apresenta um exemplo de uma rede construída considerando-se as coincidências sociais e espaço-temporais para um intervalo de encontro igual a 15 minutos.

## 5.2 Modelo Conceitual

O modelo SiST denota um grafo que é capaz de representar, exatamente, as coincidências sociais e espaço-temporais entre usuários de uma rede social online. No modelo SiST, cada nó representa um usuário e cada conexão entre dois nós indica que ambos estão envolvidos em um encontro espaço-temporal que, dependendo da existência de laços sociais na rede, pode envolver ou não os laços sociais dos usuários em questão. Em uma rede instanciada a partir do modelo SiST - chamada aqui de rede SiST - a ligação entre dois nós só existe se ambos os usuários representados por aqueles nós possuem *check-ins* no mesmo local e se a diferença temporal entre



**Figura 5.4:** Rede construída considerando-se informações sociais e espaço-temporais para  $\Delta t = 15$  min.

os referidos *check-ins* for menor ou igual ao intervalo de encontro pré-determinado. Ou seja, a ligação entre dois nós só existe se houver um encontro espaço-temporal no mundo físico.

Ao contrário das redes da Figura 5.3 e da Figura 5.4, as redes SiST não representam as várias situações de encontro por meio de múltiplas conexões entre os nós, mas sim por meio de um atributo nas conexões que é uma série temporal que indica a presença de vários instantes de tempo e vários locais de encontro. A unidade temporal da referida série é o instante de tempo do *check-in* precursor participante do encontro. O modelo foi definido desta forma com o objetivo de apresentar menor custo de armazenamento e otimizar o processamento das redes SiST.

Uma rede SiST  $G = (N, E, \Delta t, TS)$  é um grafo que consiste de um conjunto finito de  $N$  nós e  $E$  conexões. Dois nós  $a$  e  $b$  de um conjunto  $N$  representam dois usuários  $u_a$  e  $u_b$ . A conexão  $e$  entre os nós  $a$  e  $b$  existe, se e somente se,  $u_a$  e  $u_b$  são participantes de um encontro espaço-temporal (social ou não, a depender do tipo de encontro modelado) e, finalmente, se o intervalo de tempo entre os *check-ins* de  $u_a$  e  $u_b$  respeitam o intervalo de encontro  $\Delta t$ . A conexão  $e$  possui uma série temporal ( $ts_{ab}$ ) que representa os diversos encontros detectados entre os usuários  $a$  e  $b$ . Formalmente, uma rede SiST é apresentada na Definição 4.

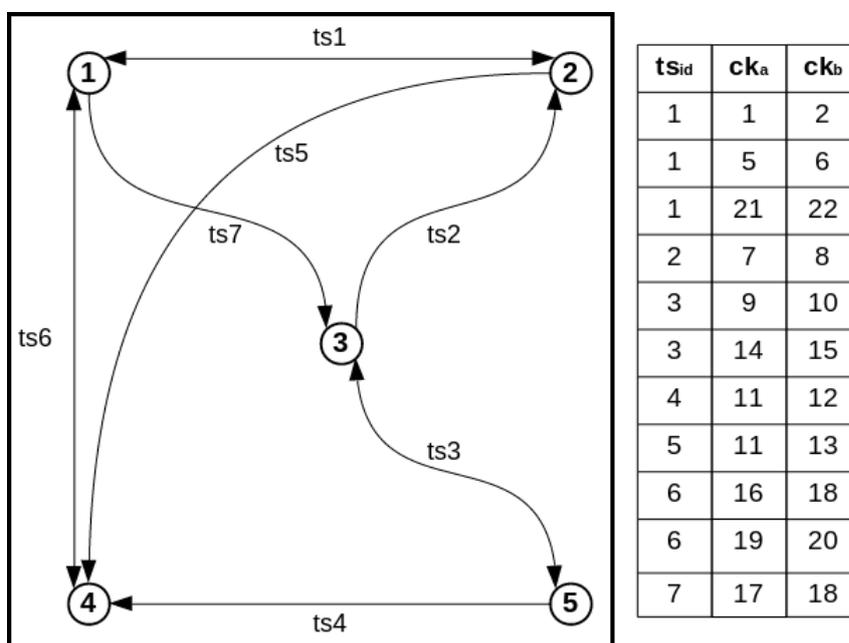
**Definição 4.** Seja  $G = (N, E, \Delta t, TS)$ , onde  $N$  é o conjunto de nós,  $E$  é o conjunto de conexões,  $\Delta t$  é o intervalo de encontro,  $TS$  é o conjunto de mapeamentos para as  $E$  séries temporais existentes no grafo. Dois nós  $a$  e  $b$ , ambos contidos em  $N$ , representando os usuários  $u_a$  e  $u_b$ , respectivamente, são relacionados por uma conexão  $e$ , contida em  $E$ , se e somente se  $a$  e  $b$  participam de, pelo menos, um encontro definido pela Definição 1.

Para cada par de nós  $a$  e  $b$  conectados, existe uma série temporal  $ts_{ab}$  de tamanho igual a  $M$ , sendo  $M$  o número de encontros existentes entre os usuários  $u_a$  e  $u_b$ . Assim sendo, pode-se definir uma série temporal de uma conexão do modelo SiST de acordo com a Definição 5.

**Definição 5.**  $ts_{ab} = \{(ck_{a_1}, ck_{b_1}), \dots, (ck_{a_i}, ck_{b_i})\}$ , onde  $ck_{a_1}$  e  $ck_{b_1}$  é o par de *check-ins* que representa o primeiro encontro entre os usuários  $u_a$  e  $u_b$  e  $ck_{a_i}$  e  $ck_{b_i}$  representa o par de *check-ins* referente ao  $i$ -ésimo encontro entre ambos usuários.

A conexão entre dois usuários  $u_a$  e  $u_b$  é representada, formalmente, pela notação a seguir:  $((u_a)\{CK_a\}) \rightarrow ((u_b)\{CK_b\})$ , onde  $u_a$  é o usuário seguidor,  $u_b$  é o usuário precursor,  $\{CK_a\}$  e  $\{CK_b\}$  são os conjuntos de *check-ins* de  $u_a$  e  $u_b$ , respectivamente, que fazem parte da série temporal  $ts_{ab}$ . Assim, se entre dois usuários  $a$  e  $b$  ocorrem 3 encontros, a conexão entre ambos é representada por  $((a)\{ck_{a_1}; ck_{a_2}; ck_{a_3}\}) \rightarrow ((b)\{ck_{b_1}; ck_{b_2}; ck_{b_3}\})$ , sendo cada encontro entre  $a$  e  $b$  representado por  $(ck_{a_1} \rightarrow ck_{b_1})$ ,  $(ck_{a_2} \rightarrow ck_{b_2})$  e  $(ck_{a_3} \rightarrow ck_{b_3})$ .

A Figura 5.5 apresenta, ao lado esquerdo, uma rede SiST de acordo com as definições previamente apresentadas, modelando encontros sociais e gerada a partir do cenário exposto na Seção 5.1.1. Neste caso, os encontros sociais são representados considerando-se  $\Delta t$  igual a 30 minutos. Cada conexão da referida rede apresenta um identificador para sua série temporal. Cada série é mostrada na tabela ao lado direito da Figura 5.5. Por exemplo, pode-se perceber que entre os nós 1 e 2 a conexão possui o rótulo  $ts1$  que é um apontador para a série temporal desta conexão. Ao se analisar a tabela do lado direito da referida figura, nota-se que a série temporal de identificador igual a 1 possui três pares de *check-ins*, são eles: (1,2), (5,6) e (21,22).



**Figura 5.5:** Exemplo de Rede SiST baseada no cenário hipotético da Seção 5.1.1.

Diante disso, na tentativa de generalizar o modelo proposto, as informações presentes no mesmo dependem, exclusivamente, das informações contidas no *check-in* publicado. Por exemplo, no Twitter o usuário pode realizar um *check-in* e, agregado àquele *check-in*, o mesmo pode realizar algum comentário relacionado ao *check-in*. A última coluna da Tabela 5.1 contém

exemplos destes comentários. Deste modo, em cada unidade da série temporal tem-se informações a respeito do usuário participante do encontro (*uid*), o local em que ocorreu o encontro (*pid*), o instante de tempo em que se deu a publicação do encontro (*timestamp*) e, por fim, informações textuais sobre a publicação do encontro (*texto*).

O modelo conceitual é exibido na Figura 5.6 por meio de um diagrama de classes em UML. Este diagrama relaciona, portanto, os conceitos do modelo SiST com os conceitos de redes sociais equipadas com a funcionalidade de *check-in*.

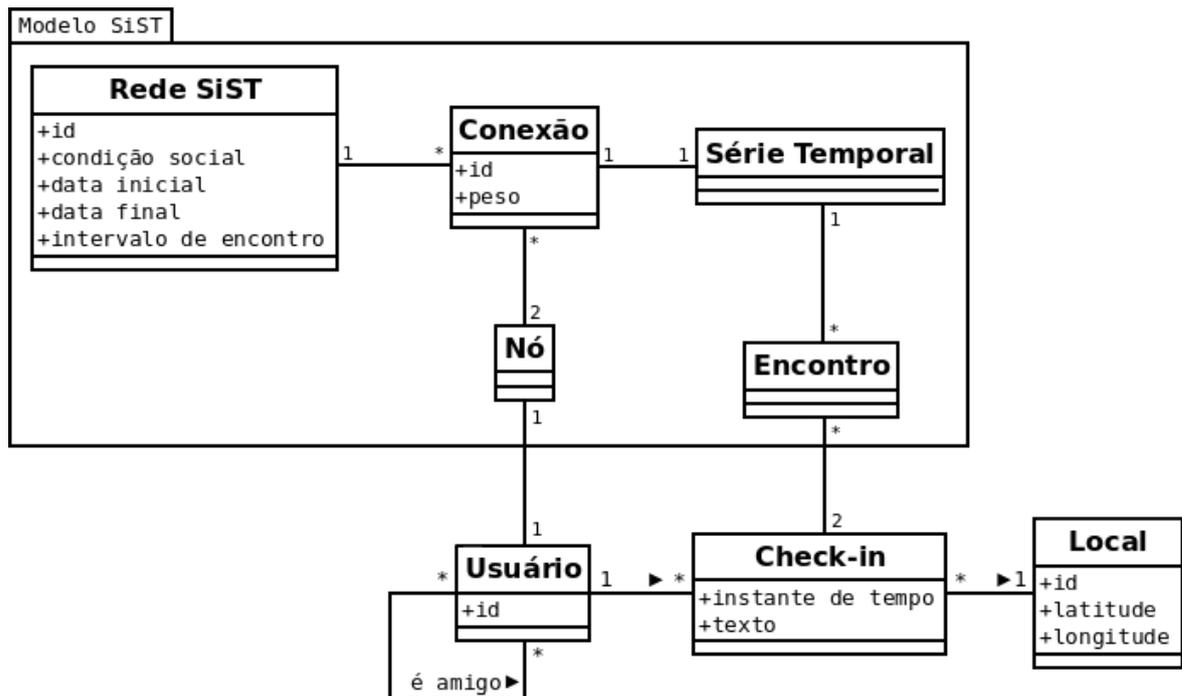


Figura 5.6: Modelo conceitual representado por um Diagrama de Classes em UML.

O diagrama da Figura 5.6 mostra que o modelo SiST é composto por cinco classes conceituais: a Rede SiST, a conexão, o nó, a série temporal e o encontro. No diagrama, essas classes são relacionadas com outras classes do domínio do problema: o usuário, o *check-in* e o local. Nota-se, portanto, que uma rede SiST é composta por muitas conexões; cada conexão conecta dois nós e possui uma série temporal. Uma série temporal, por sua vez, é composta por vários encontros cada qual sendo caracterizado por possuir dois *check-ins*; cada *check-in* é publicado em um local e por um usuário e, finalmente, um usuário é amigo de vários usuários.

### 5.3 Modelo Lógico de Dados

O modelo lógico implementa as cinco classes conceituais apresentadas na Seção 5.2: encontro, nó, conexão, série temporal e rede SiST. Essas classes são, também, as classes lógicas do modelo e foram implementadas em Java<sup>1</sup>. Além disso, a classe representativa de um *check-in*

<sup>1</sup><http://www.oracle.com/java/index.html>

é apresentada para que o modelo seja melhor compreendido.

Do mesmo modo, foi projetado um esquema relacional gerado a partir do modelo conceitual apresentado anteriormente. O esquema relacional foi implementado no PostgreSQL<sup>2</sup>, um Sistema Gerenciador de Banco de Dados que oferece amplo suporte a dados geográficos através do módulo PostGIS<sup>3</sup>. O objetivo desta seção é mostrar o modelo lógico de dados que define como o modelo conceitual foi implementado. Estas classes são apresentadas considerando-se apenas os atributos mais importantes. As operações são descritas nas seções posteriores.

Como já apresentado no Capítulo 4, um *check-in* é basicamente composto por um identificador do *check-in*, um identificador do usuário que efetuou aquele *check-in*, um identificador do local em que o *check-in* ocorreu e, finalmente, pela informação sobre o instante de tempo do *check-in*. Além disto, um *check-in* pode, opcionalmente, ser composto por um texto. Um *check-in* é representado por meio da classe `Checkin`.

```
public class Checkin {
    private int id;
    private int userId, placeId;
    private Timestamp timeInstant;
    private String text;

    //métodos getters e setters
}
```

A classe `Encounter` representa um encontro formado por um par de *check-ins*: o *check-in* que foi realizado primeiro (`targetCheckin`), publicado pelo usuário *precursor* e o *check-in* que foi realizado depois (`sourceCheckin`), publicado pelo usuário *seguidor*.

```
public class Encounter {
    private Checkin sourceCheckin, targetCheckin;

    //métodos getters e setters
}
```

A rede SiST é um grafo. Este grafo é composto por nós e conexões. Um nó é implementado pela classe `Node` que, por sua vez, representa um, e somente um, usuário da rede social que é identificado por um identificador único. A classe `Edge` representa uma conexão direcionada entre dois nós e é composta por uma lista de encontros. Esta lista de encontros é, na verdade, a série temporal da referida conexão.

---

<sup>2</sup><http://www.postgresql.org/>

<sup>3</sup><http://postgis.net/>

```
public class Node {
    private int userId;

    //métodos getters e setters
}

public class Edge {
    private Node sourceNode, targetNode;
    private List<Encounter> encountersTimeSeries;

    //métodos getters e setters
}
```

Finalmente, a classe `SiSTNetwork` representa a rede SiST tal qual é proposta neste trabalho. Além dos nós (atributo `nodes`) e conexões (atributo `edges`), esta classe é composta, também, por um atributo identificador da rede (`id`), pelo atributo `socialCondition`, que indica se a rede é uma rede de encontros sociais ou não sociais, por `initialDate` e `finalDate`, que representam o período em que os encontros representados na rede foram detectados e, finalmente, por `encounterIntervalSec`, que indica o intervalo de tempo dos encontros representados na rede em questão (i.e. o intervalo de encontro, em segundos).

```
public class SiSTNetwork{

    private int id;
    private boolean socialCondition;
    private Timestamp initialDate, finalDate;
    private int encounterIntervalSec;
    private List<Encounter>[][] matrix;
    private List<Node> nodes;
    private List<Edge> edges;

    public void loadAdjacencyMatrix(){ //... }

    public List<Encounter> getAllEncounters(){
        //... }

    public List<Encounter> getAllEncounters(int userId){
        //... }

    public List<Encounter> getAllEncounters(Timestamp
```

```
        initialDate, Timestamp finalDate){ //... }

public List<List<Encounter>> getDuplicatedEncounters() {
    //... }

public List<GroupEncounter> getGroupEncounters() {
    //... }

public List<LeadershipPattern> getLeadershipPattern() {
    //... }

//métodos getters e setters
}
```

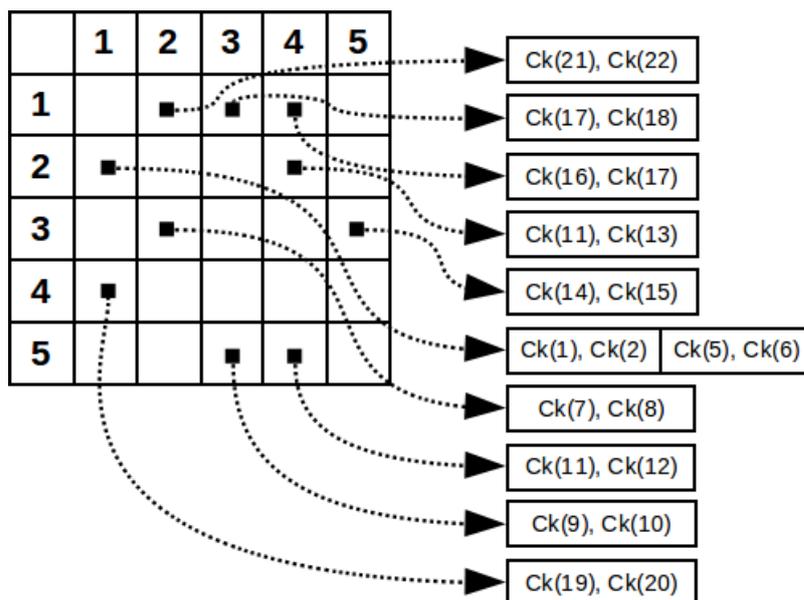
O atributo `matrix` representa a matriz de adjacência que reproduz a rede em questão. Esta matriz faz parte do modelo físico de dados e é apresentada na Seção 5.4, a seguir. Os métodos restantes, apresentados na classe `SiSTNetwork`, são descritos na Seção 5.5

#### 5.4 Modelo Físico de Dados

A matriz de adjacência do modelo SiST é denominada matriz de adjacência espaço-temporal. Essa matriz é caracterizada por também representar a série temporal de cada conexão do grafo. Desta forma, cada elemento  $a_{ij}$  da matriz de adjacência espaço-temporal denota uma conexão direcionada entre os nós  $i$  e  $j$ , onde elemento  $a_{ij}$  é caracterizado por possuir um vetor de ponteiros que representa a série temporal da conexão. A Figura 5.7 mostra a referida matriz para a rede SiST gerada a partir do cenário hipotético da Seção 5.1.

A matriz de adjacência da Figura 5.7 representa, portanto, todos os encontros presentes na rede SiST da Figura 5.5. Tome-se como exemplo a conexão  $ts_1$  da rede apresentada na Figura 5.5. Como a referida conexão é bidirecional, conectando os nós 1 e 2, isso indica que há situações de encontros em que o usuário representado pelo nó 1 é o precursor e, por outro lado, também existem situações em que o usuário representado pelo nó 2 é o precursor. Assim, as células da matriz de adjacência espaço-temporal que representa ambas situações são aquelas em que relacionam os nós 1 e 2 (i.e. a célula  $a_{21}$  representa os encontros em que o nó 1 foi o precursor e a célula  $a_{12}$  representa os encontros em que o nó 2 foi o precursor).

Uma conexão unidirecional é representada apenas por meio da descrição da série temporal na coluna do nó precursor. Este fato pode ser observado para a conexão  $ts_2$  da rede apresentada na Figura 5.5, em que conecta os nós 2 e 3, por exemplo. Nesta situação, pode-se perceber que como o usuário 2 é o precursor, a série temporal da referida conexão é representada por meio da  $a_{32}$ , em que o usuário 2 é representado na coluna da matriz.



**Figura 5.7:** Exemplo de matriz de adjacência espaço-temporal baseado no cenário hipotético da Seção 5.1.1.

A matriz de adjacência é gerada no método `loadAdjacencyMatrix` da classe `SiSTNetwork`. Essa matriz é de fundamental importância pois é por meio dela que a rede é manipulada. Por exemplo, o método `getAllEncounters()` e suas variantes fazem uso dessa matriz para retornar os encontros representados pela rede em questão. Além disso, a matriz é manipulada, também, na detecção de padrões de mobilidade de usuários definidos aqui, como o padrão de grupos de encontros (executado pelo método `getGroupEncounters()`) e o padrão de liderança de encontros (executado pelo método `getLeadershipPattern()`). O padrão de encontros duplicados (executado pelo método `getDuplicatedEncounters()`) é um padrão de publicação de encontros que todos os padrões de mobilidade previamente citados fazem uso. Os algoritmos de detecção desses padrões bem como o algoritmo de criação de redes SiST são discutidos na Seção 5.5, descrita a seguir.

## 5.5 Algoritmos que compõem o Modelo SiST

A Seção 5.5.1 descreve o algoritmo de construção de uma rede SiST. A Seção 5.5.2 define o padrão de *encontros duplicados* bem como apresenta o algoritmo implementado para sua detecção. Na Seção 5.5.3, o padrão *grupo de encontros* é definido e, da mesma forma, o algoritmo que detecta grupos de encontros é apresentado e, finalmente, a Seção 5.5.4 descreve o padrão de *liderança de encontros* e exhibe o algoritmo para detecção desse padrão.

### 5.5.1 Algoritmo de Construção de Redes SiST

O processo de construção de uma rede SiST é implementado pela descrição apresentada no Algoritmo 2. Este algoritmo é demonstrado também no diagrama de atividades da Figura

5.8. Primeiramente, é inicializado o grafo vazio em que é construída a rede SiST (linha 1). O Algoritmo 1 é, então, executado dentro do Algoritmo 2 e todos os encontros detectados pelo mesmo são atribuídos à variável *encontros* (linha 2). O algoritmo inicia a construção do grafo por meio do laço de repetição entre as linhas 3 e 12 que itera sobre cada encontro *enc* de *encontros*.

---

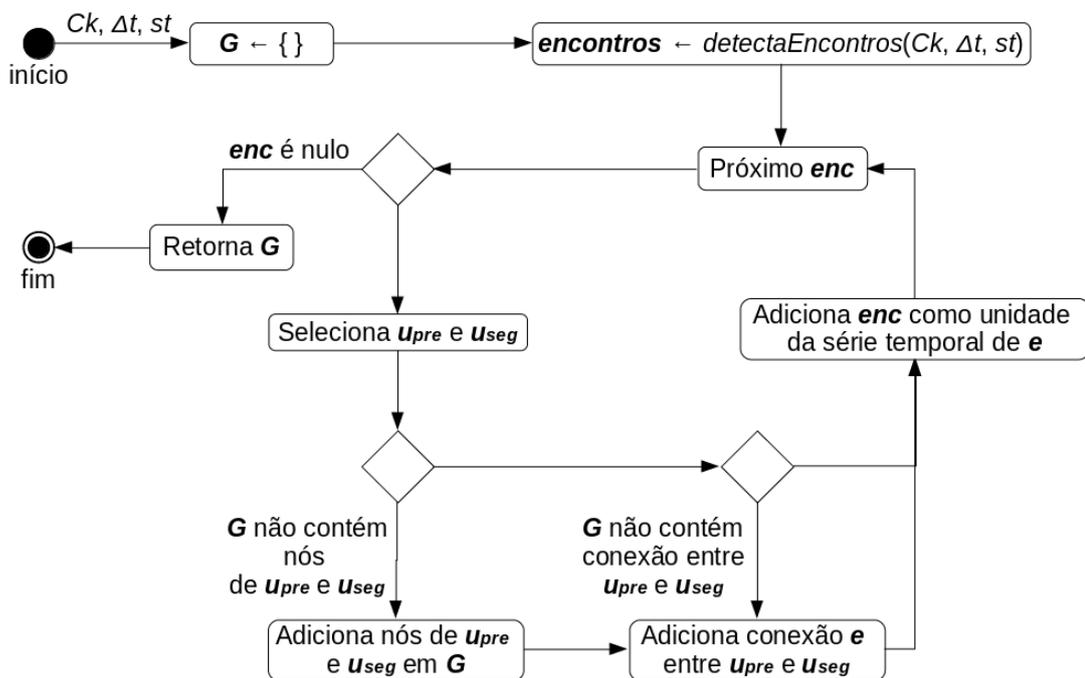
**Algoritmo 2** Construtor\_SiST( $Ck, \Delta t, st$ )
 

---

```

1:  $G \leftarrow \{\}$ ; // grafo direcional a ser retornado (inicialmente vazio)
2:  $encontros \leftarrow detectaEncontros(Ck, \Delta t, st)$ ; // executa o Algoritmo 1
3: PARA cada encontro enc de encontros:
4:   SELECIONA o usuário precursor ( $u_{pre}$ ) e o usuário seguidor ( $u_{seg}$ ) de enc;
5:   SE  $G$  não contém os nós de  $u_{pre}$  e  $u_{seg}$  ENTÃO
6:     ADICIONA os nós de  $u_{pre}$  e  $u_{seg}$ ;
7:   FIM SE
8:   SE não existe conexão  $e$  entre  $u_{pre}$  e  $u_{seg}$  ENTÃO
9:     ADICIONA conexão  $e$  entre  $u_{pre}$  e  $u_{seg}$ ;
10:  FIM SE
11:  ADICIONA o par de check-ins de enc como unidade da série temporal  $ts$  de  $e$ ;
12: FIM PARA
13: RETORNE  $G$ ;
  
```

---

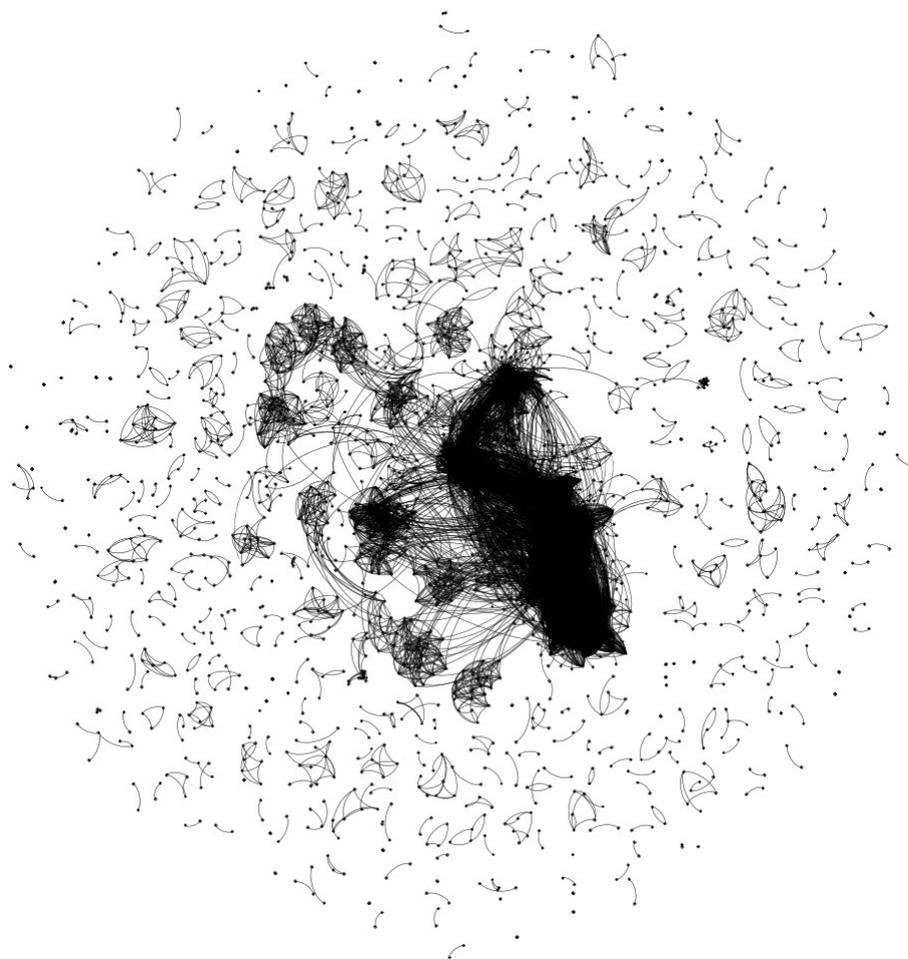


**Figura 5.8:** Diagrama de atividades do Algoritmo 2.

A descrição na linha 4 do Algoritmo 2 indica que em cada iteração do laço, os usuários de *enc* são recuperados para que a construção do grafo seja realizada nas linhas seguintes do algoritmo. A partir daí, a estrutura condicional delimitada entre as linhas 5 e 7 verifica se o grafo  $G$  a ser retornado já possui os nós representantes dos referidos usuários. Caso não possua, esses

nós são adicionados ao grafo (linha 6). Na estrutura condicional delimitada entre as linhas 8 e 10, o algoritmo verifica se já existe conexão entre os nós envolvidos no encontro analisado na iteração corrente do laço. Caso essa conexão não exista, a mesma é criada (linha 9). Por fim, para cada encontro *enc*, a série temporal da conexão entre os nós envolvidos no encontro é incrementada, i.e. *enc* é adicionado como unidade da série temporal de *e* (linha 11).

Os grafos (i.e. as redes SiST) gerados a partir do Algoritmo 2 são persistidos em uma base de dados projetada tal qual o modelo conceitual apresentado na Seção 5.2. Como normalmente as redes SiST possuem uma grande quantidade de nós e conexões, as mesmas dem ser carregadas e analisadas em programas de análise de redes complexas, como o Gephi<sup>4</sup> ou o NetworkX<sup>5</sup>, por exemplo. A Figura 5.9 mostra um exemplo real de uma rede SiST construída a partir do Algoritmo 2 e da base GOW\_SCE. Nota-se que embora a rede possua um alto número de nós e conexões (neste caso, 3.252 nós e 16.235 conexões), a mesma possui muitos nós desconectados. No processo de validação do modelo SiST, pode-se observar que essa característica é comum para as redes construídas a partir do mesmo, i.e. redes SiST possuem baixo grau de conectividade.



**Figura 5.9:** Rede SiST construída a partir da base GOW\_SCE considerando  $\Delta t = 15$  minutos.

<sup>4</sup><https://gephi.org/>

<sup>5</sup><https://networkx.github.io/>

Finalmente, como cada encontro é representado como uma unidade de uma série temporal, caso um ou dois usuários republiquem, propositadamente ou não, um ou mais encontros no mesmo local, esses encontros são representados independentemente dessa duplicidade. A consideração sobre se um encontro foi duplicado ou não é dependente de aplicação e, por isso, esse evento é definido aqui como um padrão de publicação de encontros: os encontros duplicados.

### 5.5.2 Algoritmo de Detecção de Encontros Duplicados

Em dados reais de *check-ins*, pode-se observar que há situações de usuários que republi- cam *check-ins*. Por exemplo, considere o conjunto de *check-ins* da Tabela 5.3. Nota-se que o usuário 13 republica o seu *check-in* no local 200 após cerca de um minuto. Do mesmo modo, pode-se observar que usuário 19 faz o mesmo por duas vezes: na primeira, o mesmo realiza um *check-in* em cerca de três minutos depois do primeiro e na sequência publica um *check-in* no mesmo local duas horas depois do primeiro *check-in*.

$ck_{id}$	$user_{id}$	$place_{id}$	<i>timestamp</i>
1	13	200	18/03/2009 15:19:47
2	13	200	18/03/2009 15:20:41
3	19	200	18/03/2009 15:21:54
4	19	200	18/03/2009 15:23:14
5	19	200	18/03/2009 17:20:50

**Tabela 5.3:** Exemplo de *check-ins* publicados repetidamente.

As redes SiST são construídas independentemente dessa repetição. Ou seja, uma vez que não tem como saber as razões pelas quais ocorrem essas repetições de *check-ins*, preferiu-se que as mesmas fossem consideradas como parte da série temporal das conexões presentes nas redes. Assim, considerando que os *check-ins* da Tabela 5.3 fazem parte de um conjunto de *check-ins* de uma rede social real, uma rede SiST construída por meio do Algoritmo 2 e um intervalo de encontro de 15 minutos apresentaria uma conexão entre os nós 13 e 19 que seria representada por  $((19)\{3;3;4;4\}) \rightarrow ((13)\{1;2;1;2\})$ .

Desta forma, pode-se notar que o que aparenta ser uma situação de encontro (considerando-se o intervalo de encontro de 15 minutos), é representado pela rede SiST como sendo quatro situações de encontros diferentes, a saber  $(3 \rightarrow 1)$ ,  $(3 \rightarrow 2)$ ,  $(4 \rightarrow 1)$  e  $(4 \rightarrow 2)$ . A partir disso, define-se um *encontro duplicado* conforme indicado pela Definição 6.

**Definição 6. Encontro Duplicado:** O conjunto de  $i$  encontros  $\{(ck_a \rightarrow ck_b)_1; \dots; (ck_a \rightarrow ck_b)_i\}$  entre os usuários  $a$  e  $b$  representam um encontro duplicado se  $ck_b(time)_i - ck_b(time)_1 \leq \Delta t$ .

O Algoritmo 3 descreve como se deu a implementação da detecção de encontros duplicados em uma rede SiST. A representação gráfica deste algoritmo é apresentada na Figura 5.10 por meio de um diagrama de atividades. O algoritmo recebe uma rede SiST como parâmetro  $G$  e,

para cada usuário precursor representado em  $G$ , o mesmo compara cada par de encontros em que esse usuário foi, de fato, precursor. Esse processo é realizado por meio de duas estruturas de repetição aninhadas: a primeira entre as linhas 3 e 16 e a segunda entre as linhas 6 e 15. Desta forma, para cada par de encontro do usuário iterado, o algoritmo analisa se ambos encontros ocorreram no mesmo local (linha 7), se ambos possuem o mesmo seguidor (linha 8) e, por fim, verifica se o instante de tempo dos *check-ins* precursores de ambos encontros respeitam o intervalo de tempo  $\Delta t$  (linha 10). Caso os dois pares de encontro de cada usuário analisado tenham ocorrido no mesmo local, envolvam os mesmos seguidores e tenham ocorrido em um intervalo de tempo menor do que  $\Delta t$ , então ambos encontros são considerados duplicados. Assim, esses encontros são adicionados na lista de retorno do algoritmo (linha 11).

---

**Algoritmo 3** DetectaEncontrosDuplicados( $G$ )
 

---

```

1:  $LISTA \leftarrow \{\}$ ; // lista de encontros duplicados a ser retornada (inicialmente vazia)
2: SELECIONA todos os usuários precursores de  $G$ ;
3: PARA cada usuário precursor  $u_{pre}$  de precursores:
4:   SELECIONA todos os encontros  $encs_{pre}$  em que  $u_{pre}$  é precursor;
5:   ORDENA  $encs_{pre}$  pelo instante de tempo do precursor;
6:   PARA cada par de encontro  $enc_{pre1}$  e  $enc_{pre2}$  de  $encs_{pre}$ 
7:     SE  $enc_{pre1}$  e  $enc_{pre2}$  ocorreram no mesmo local ENTÃO
8:       SE  $enc_{pre1}$  e  $enc_{pre2}$  têm o mesmo seguidor ENTÃO
9:         SELECIONA o instante de tempo  $t_1$  de  $enc_{pre1}$  e  $t_2$  de  $enc_{pre2}$ ;
10:        SE  $|t_2 - t_1| \leq \Delta t$  ENTÃO
11:          ADICIONA  $enc_{pre1}$  e  $enc_{pre2}$  em  $LISTA$ ; //são encontros duplicados;
12:        FIM SE
13:      FIM SE
14:    FIM SE
15:  FIM PARA
16: FIM PARA
17: RETORNE  $LISTA$ ;

```

---

### 5.5.3 Algoritmo de Detecção de Grupo de Encontros

Cada conexão de uma rede SiST representa um ou vários encontros ocorridos entre um par de usuários. Estes, por sua vez, são representados pelos nós da rede. A manipulação da rede, por meio de sua matriz de adjacência, permite que sejam detectados encontros ocorridos com mais do que dois usuários. Este tipo de encontro é definido aqui como um grupo de encontros.

**Definição 7. Grupo de Encontros:** Três ou mais usuários de um conjunto de  $N$  usuários são participantes de um grupo de encontros  $S$  se cada usuário de  $S$  é participante de, pelo menos, um encontro com outro usuário de  $S$ .

Observe, por exemplo, os *check-ins* da Tabela 5.4. Neste caso, considerando-se um  $\Delta t = 15$  minutos, pode-se notar que além da detecção de dois pares de encontros (i.e.  $((20)\{2\}) \rightarrow$

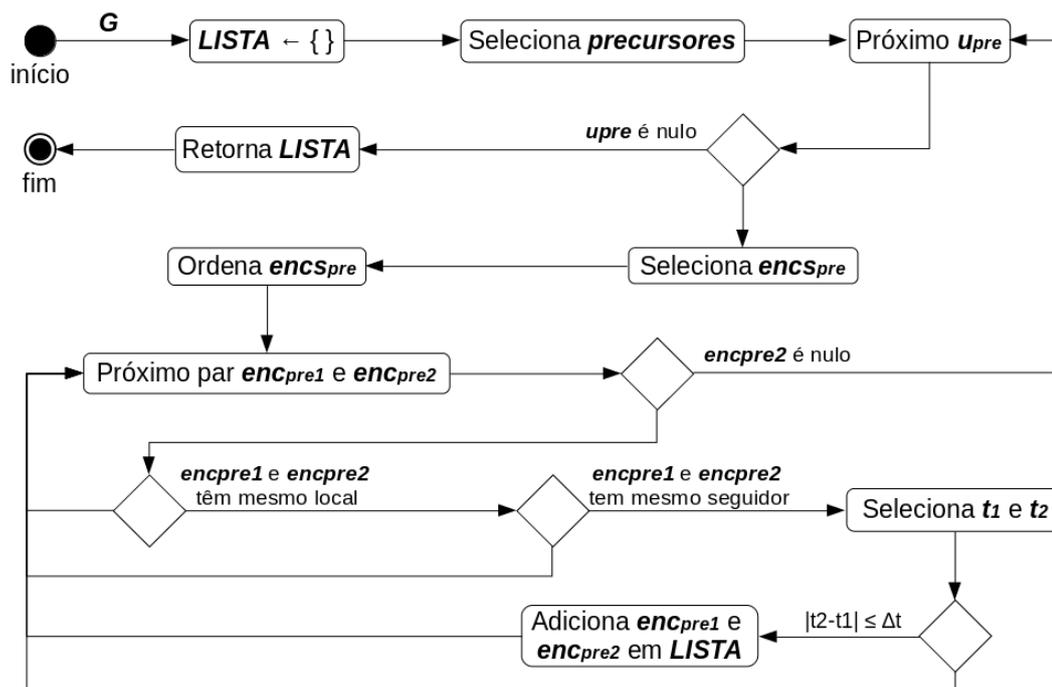


Figura 5.10: Diagrama de atividades do Algoritmo 3.

$((10)\{1\})$  e  $((10)\{9\}) \rightarrow ((70)\{8\})$ ), existe também um grupo de encontros formado pelos seguintes encontros:  $((40)\{4\}) \rightarrow ((30)\{3\})$ ,  $((50)\{5\}) \rightarrow ((30)\{3\})$ ,  $((50)\{5\}) \rightarrow ((40)\{4\})$ ,  $((40)\{6\}) \rightarrow ((50)\{5\})$  e, finalmente,  $((60)\{7\}) \rightarrow ((40)\{6\})$ . Ou seja, o referido grupo de encontros é composto pelos *check-ins* 3, 4, 5, 6 e 7 e, conseqüentemente, pelos seus usuários: 30, 40, 50 e 60. O Algoritmo 4 corresponde ao algoritmo para detecção de grupos de encontros.

$ck_{id}$	$user_{id}$	$place_{id}$	<i>timestamp</i>
1	10	300	12/03/2015 2:16:15
2	20	300	12/03/2015 2:21:45
3	30	300	12/03/2015 5:12:02
4	40	300	12/03/2015 5:19:08
5	50	300	12/03/2015 5:27:09
6	40	300	12/03/2015 5:40:04
7	60	300	12/03/2015 5:50:53
8	70	300	12/03/2015 9:28:15
9	10	300	12/03/2015 9:32:18

Tabela 5.4: Exemplo de *check-ins* participantes de grupos de encontros.

O Algoritmo 4 busca todos os grupos de encontros para cada nó precursor de um grafo  $G$  (laço que inicia-se na linha 3). O padrão de grupos de encontros faz uso dos encontros duplicados, ou seja, um grupo de encontros pode ser composto por encontros duplicados e não duplicados. Por esta razão é que na linha 5 todos os encontros duplicados e não duplicados são considerados, estes encontros são ordenados na linha 6. A estrutura de repetição entre as linhas 7 e 28, por sua vez, itera sobre cada encontro de  $encs_{pre}$  com o objetivo de identificar grupos de encontros.

**Algoritmo 4** DetectaGrupoEncontros( $G$ )

---

```

1:  $GRUPOS \leftarrow \{\}$ ; // lista de grupos de encontros a ser retornada (inicialmente vazia)
2: SELECIONA todos os usuários  $precursores$  de  $G$ ;
3: PARA cada usuário precursor  $u_{pre}$  de  $precursores$ 
4:    $grupo \leftarrow \{\}$ ; // lista que conterá todos os encontros que formam um grupo
5:   SELECIONA todos os encontros  $encs_{pre}$  em que  $u_{pre}$  é precursor (duplicados ou não);
6:   ORDENA  $encs_{pre}$  pelo instante de tempo do precursor;
7:   PARA cada encontro  $enc$  de  $encs_{pre}$ 
8:     SE  $grupo$  está vazio ENTÃO
9:       ADICIONA  $enc$  em  $grupo$ ;
10:    SENÃO
11:      SE  $enc$  e  $grupo$  ocorreram no mesmo local ENTÃO
12:        SELECIONA último encontro  $enc_{ult}$  de  $grupo$ ;
13:        SELECIONA instante de tempo  $t_{ult}$  de  $enc_{ult}$ ;
14:        SELECIONA instante de tempo  $t_{prec}$  do precursor de  $enc$ ;
15:        SE  $|t_{prec} - t_{ult}| \leq \Delta t$  ENTÃO
16:          ADICIONA  $enc$  em  $grupo$ ;
17:        SENÃO
18:          SE grupo têm mais do que 2 usuários ENTÃO
19:            ADICIONA  $grupo$  a  $GRUPOS$ ;
20:          FIM SE
21:          FIM SE-SENÃO
22:        SENÃO
23:          SE grupo têm mais do que 2 usuários ENTÃO
24:            ADICIONA  $grupo$  a  $GRUPOS$ ;
25:          FIM SE
26:          FIM SE-SENÃO
27:        FIM SE-SENÃO
28:      FIM PARA
29: FIM PARA
30: RETORNE  $GRUPOS$ ;

```

---

Na primeira execução do laço entre as linhas 7 e 28 para cada usuário precursor, o grupo é inicialmente vazio e, portanto, o primeiro encontro  $enc$  analisado é adicionado diretamente ao grupo (linha 9). A partir da segunda iteração, o algoritmo avalia, primeiramente, se o grupo sendo criado (variável  $grupo$ ) e o encontro da iteração (variável  $enc$ ) ocorreram no mesmo local (linha 11). Caso ambos tenham ocorrido no mesmo local, o algoritmo verifica se o intervalo de tempo entre o último encontro de  $grupo$  e o instante de tempo do precursor de  $enc$  respeitam  $\Delta t$  (linha 15). Em caso afirmativo, o encontro  $enc$  é adicionado à variável  $grupo$ . Caso contrário, o algoritmo verifica se o grupo em construção possui mais do que dois usuários participantes (linha 18) e, então, adiciona-o na variável de retorno  $GRUPOS$ .

O Algoritmo 4 retorna, portanto, todos os grupos de encontros publicados em uma rede social online. Este algoritmo é representado também por meio de um diagrama de atividades

tal qual demonstrado na Figura 5.11. A definição desse algoritmo está diretamente relacionada com estudos sobre mobilidade de usuários, uma vez que por meio de grupos de usuários pode-se analisar padrões de movimentação, identificar eventos, dentre outros. Além disso, a definição de grupos de encontros oferece oportunidades para elaboração de novas estratégias de marketing em redes sociais online, como por exemplo o oferecimento de promoções em estabelecimentos comerciais em que grupos de encontros são frequentemente detectados.

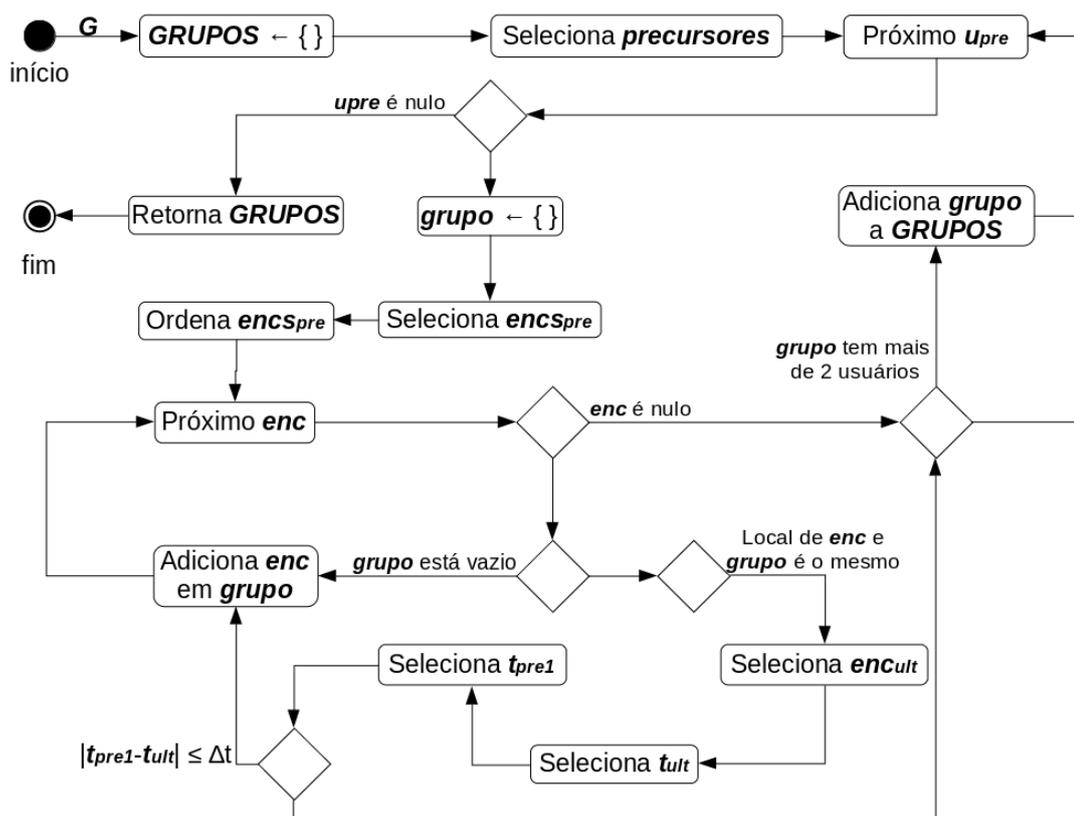


Figura 5.11: Diagrama de atividades do Algoritmo 4.

#### 5.5.4 Algoritmo de Detecção de Liderança de Encontros

Ao se analisar os encontros, foi possível observar que alguns usuários publicam encontros com outros usuários repetidamente, independentemente do local. Além disso, pode-se notar também que normalmente esses encontros são liderados pelo mesmo usuário. Ou seja, na ocorrência desses encontros, o usuário precursor é frequentemente o mesmo. Esta tese caracteriza esse evento como um padrão: o padrão de liderança de encontros. Este padrão é apresentado na Definição 8. O padrão de liderança de encontros é detectado quando um ou mais usuários são seguidores em, pelo menos,  $\theta$  situações de encontros, sejam eles em locais diferente ou não.

**Definição 8. Liderança de Encontro:** Existe liderança de um usuário  $u$  (i.e. líder) sobre  $\pi$  usuários se, e somente se,  $u$  e todos os  $\pi$  usuários se encontraram pelo menos  $\theta$  vezes e

se em todos os  $\theta$  encontros o usuário  $u$  foi o usuário precursor e os  $\pi$  usuários foram os usuários seguidores, sendo  $\theta \geq 2$ ;

O objetivo de se definir o padrão de liderança é o de auxiliar no processo de identificação de usuários influentes em uma rede social. As situações de liderança de encontros podem ser úteis considerando-se outras informações contextuais, como por exemplo informações do perfil do usuário ou dos locais em que o mesmo é “líder”. Alguns usuários, por exemplo, podem ter maior poder de influência em determinado tipo de local do que outros usuários. Diante disso, a interpretação dos resultados a partir do padrão de liderança deve ser feita considerando-se outros aspectos do contexto em que os usuários estão inseridos. Aqui é proposto, apenas, a definição do padrão bem como o algoritmo para sua detecção.

Considere, por exemplo, os *check-ins* da Tabela 5.5. Para o  $\Delta t = 15$  minutos e  $\theta = 2$ , pode-se observar que duas instâncias do padrão de liderança foram detectadas: (a) composta pelos encontros  $((15)\{2\}) \rightarrow ((36)\{1\})$ ,  $((15)\{4\}) \rightarrow ((36)\{3\})$  e  $((15)\{6\}) \rightarrow ((36)\{5\})$ ; (b) composta pelos encontros  $((11)\{8\}) \rightarrow ((36)\{7\})$  e  $((11)\{10\}) \rightarrow ((36)\{9\})$ . A observação da repetição desse padrão pode ser útil quando analisadas outras informações a respeito dos usuários envolvidos no padrão bem como dos locais em que os encontros ocorreram.

$ck_{id}$	$user_{id}$	$place_{id}$	<i>timestamp</i>
1	36	111	23/10/2015 4:23:18
2	15	111	23/10/2015 4:27:00
3	36	111	23/10/2015 6:35:35
4	15	111	23/10/2015 6:40:38
5	36	168	23/10/2015 7:27:35
6	15	168	23/10/2015 7:30:18
7	36	736	24/10/2015 6:31:31
8	11	736	24/10/2015 6:33:20
9	36	111	24/10/2015 7:38:38
10	11	111	24/10/2015 7:50:38

**Tabela 5.5:** Exemplo de *check-ins* participantes de liderança de encontros.

No caso de (a), nota-se que dos três encontros considerados, os dois primeiros pares ocorreram no mesmo local (i.e. de identificador 111). Esses pares não foram considerados como um encontro duplicado porque a diferença entre os instantes de tempo dos usuários precursores de ambos encontros é maior do que o intervalo de encontro, como pode-se observar na Tabela 5.5. O outro par de encontro participante da instância (a), por sua vez, ocorreu no local de identificador 168. Em relação a (b), pode-se observar que o referido padrão de liderança envolveu encontros ocorridos em dois locais diferentes e em um intervalo de tempo de aproximadamente uma hora. O padrão de liderança de encontro foi implementado por meio do Algoritmo 5. A Figura 5.12 representa o mesmo por meio de um diagrama de atividades.

O funcionamento do Algoritmo 5 é muito semelhante ao do Algoritmo 4, que detecta grupo de encontros. O algoritmo recebe como parâmetro uma rede SiST  $G$  e um valor para  $\theta$ .

**Algoritmo 5** DetectaLiderancaEncontros( $G, \theta$ )

---

```

1:  $INSTANCIAS \leftarrow \{\}$ ; // lista de instâncias de liderança de encontros a ser retornada
2: SELECIONA todos os usuários  $precursores$  de  $G$ ;
3: PARA cada usuário precursor  $u_{pre}$  de  $precursores$ 
4:    $instancia \leftarrow \{\}$ ; // conterà todos os encontros que formam uma instância do padrão
5:   SELECIONA todos os encontros  $encs_{pre}$  em que  $u_{pre}$  é precursor (duplicados ou não);
6:   ORDENA  $encs_{pre}$  pelo instante de tempo do precursor;
7:   PARA cada encontro  $enc$  de  $encs_{pre}$ 
8:     SE  $instancia$  está vazia ENTÃO
9:       ADICIONA  $enc$  em  $instancia$ ;
10:    SENÃO
11:      SELECIONA último encontro  $enc_{ult}$  de  $instancia$ ;
12:      SELECIONA instante de tempo  $t_{ult}$  do seguidor de  $enc_{ult}$ ;
13:      SELECIONA instante de tempo  $t_{prec}$  do precursor de  $enc$ ;
14:      SE  $|t_{prec} - t_{ult}| > \Delta t$  ENTÃO
15:        SE usuários seguidores de  $enc$  e  $enc_{ult}$  são os mesmos ENTÃO
16:          ADICIONA  $enc$  em  $instancia$ ;
17:        SENÃO
18:          SE  $instancia$  possui mais do que  $\theta$  encontros ENTÃO
19:            ADICIONA  $instancia$  a  $INSTANCIAS$ ;
20:             $instancia \leftarrow \{\}$ ;
21:          FIM SE
22:        FIM SE-SENÃO
23:      FIM SE
24:    FIM SE-SENÃO
25:  FIM PARA
26: FIM PARA
27: RETORNE  $GRUPOS$ ;

```

---

Desta forma, o primeiro laço, entre as linhas 3 e 25, itera sobre todos os nós que representam usuários precursores de  $G$ . A partir desses nós, o algoritmo investiga se existe reincidência de encontros com os mesmos seguidores. Para cada nó, este processo ocorre por meio da detecção de todos os encontros duplicados ou não (linha 5).

Assim, cada encontro, duplicado ou não, é processado por meio de outro laço de repetição, entre as linhas 7 e 24. Quando o laço mais interno está sendo executado pela primeira vez para o nó, ou seja, quando o primeiro encontro de um usuário precursor está sendo analisado, esse encontro é imediatamente adicionado à variável  $instancia$  (linha 9). A variável  $instancia$  representa uma instância do padrão de liderança que se está tentando identificar.

Se o laço mais interno já tiver realizado a primeira iteração para o nó em análise, então o algoritmo avalia se a diferença entre o instante de tempo do usuário precursor do encontro ( $t_{prec}$ ) e o instante de tempo do último encontro de  $instancia$  ( $t_{ult}$ ) é maior do que  $\Delta t$  (linha 14). Em caso afirmativo, o algoritmo verifica se os seguidores do encontro iterado ( $enc$ ) e do último encontro de  $instancia$  são os mesmos (linha 15) para, então, adicionar  $enc$  em  $instancia$  (linha 16). Na

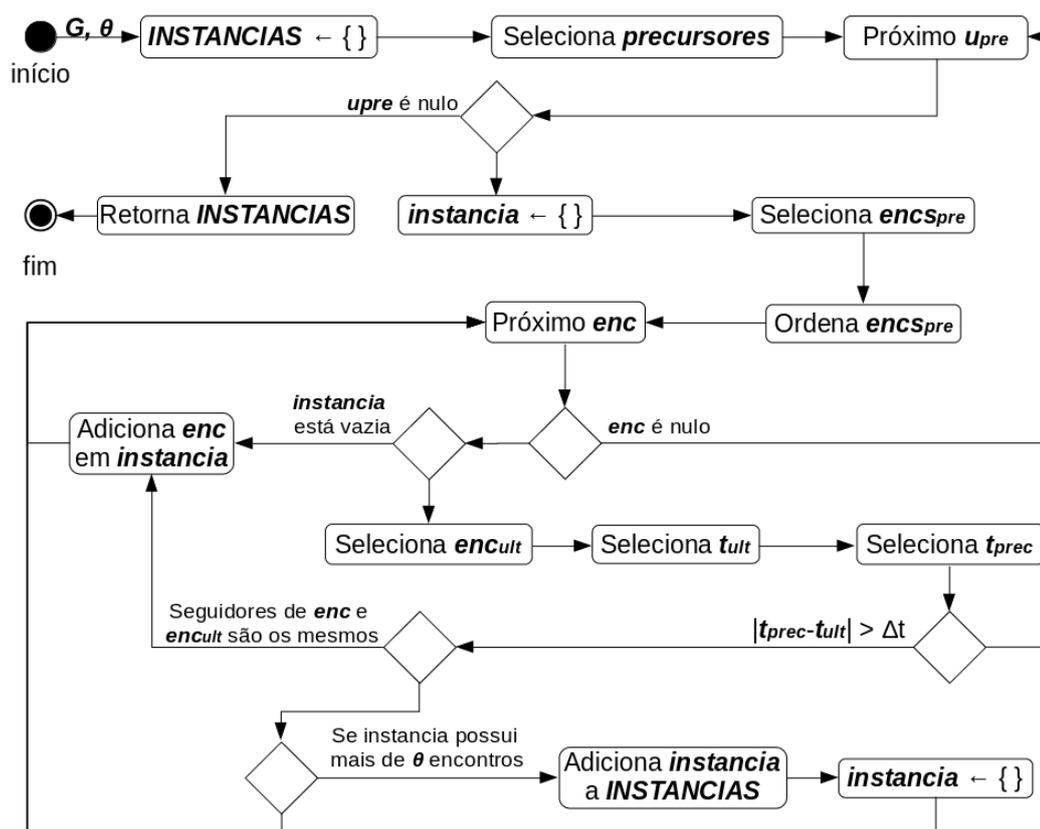


Figura 5.12: Diagrama de atividades do Algoritmo 5.

prática, essa verificação identifica se os seguidores da instância do padrão sendo detectado são os mesmos do encontro da iteração corrente (i.e. *enc*). Caso os seguidores não sejam os mesmos, então *enc* não faz parte do padrão sendo construído até o momento e, portanto, o algoritmo pode finalizar a identificação da instância corrente (i.e. variável *instancia*). Para isso, é verificado se essa instância do padrão de liderança contém, de fato, mais do que  $\theta$  encontros (linha 18) para, então, adicionar a instância na lista a ser retornada pelo algoritmo (linha 19).

Assim como o padrão de grupos de encontros, o padrão de liderança representa usuários que se locomovem no globo. Por isso, pode-se fazer uso desses padrões para definir novos serviços baseados em localização, bem como se analisar a mobilidade desses usuários. Por exemplo, pode-se investigar por grupos de encontros e lideranças de encontros com o objetivo de se identificar padrões universais de locomoção de usuários, como a identificação de locomoção de pessoas para os seus respectivos locais de trabalho ou para locais de prática de esportes. Esse padrão, assim como o modelo SiST como um todo, foram validados por meio de experimentos com dados reais de redes sociais online que são descritos na Seção 5.6.

## 5.6 Validação das Redes SiST

O modelo SiST foi avaliado por meio da execução de experimentos onde redes SiST foram criadas utilizando-se os dados descritos na Seção 4.4. Estas redes foram construídas para

quatro intervalos de encontros seguintes: 15 minutos, 30 minutos, 1 hora e 2 horas e, quando disponíveis, os laços sociais também foram considerados.

Esta seção está organizada como se segue. Na Seção 5.6.1, estas redes são analisadas sob diferentes pontos de vistas, seja por meio de uma análise estatística, seja através de uma análise temporal dos encontros que elas representam. A Seção 5.6.2, por sua vez, descreve os resultados provenientes da detecção dos padrões de mobilidade definidos neste capítulo.

### 5.6.1 Análise das Redes SiST

A análise das redes SiST foi realizada de acordo com três abordagens diferentes. Primeiro, a Seção 5.6.1.1 apresenta uma descrição das propriedades das redes, como o grau médio das redes e o coeficiente de agrupamento, por exemplo. Depois, na Seção 5.6.1.2, é descrita como se deu a análise das propriedades das redes com o objetivo de identificar a lei que governa os encontros tal qual realizado no Capítulo 4.

#### 5.6.1.1 Propriedades das Redes SiST

Para analisar as propriedades de redes SiST, o Algoritmo 2 foi executado passando-se como parâmetro um conjunto de *check-ins* ocorridos durante uma semana (i.e. a semana com maior quantidade de publicações para cada base de dados). A análise das propriedades se deu por cada dia da semana e os resultados foram comparados com aqueles descritos no Capítulo 4.

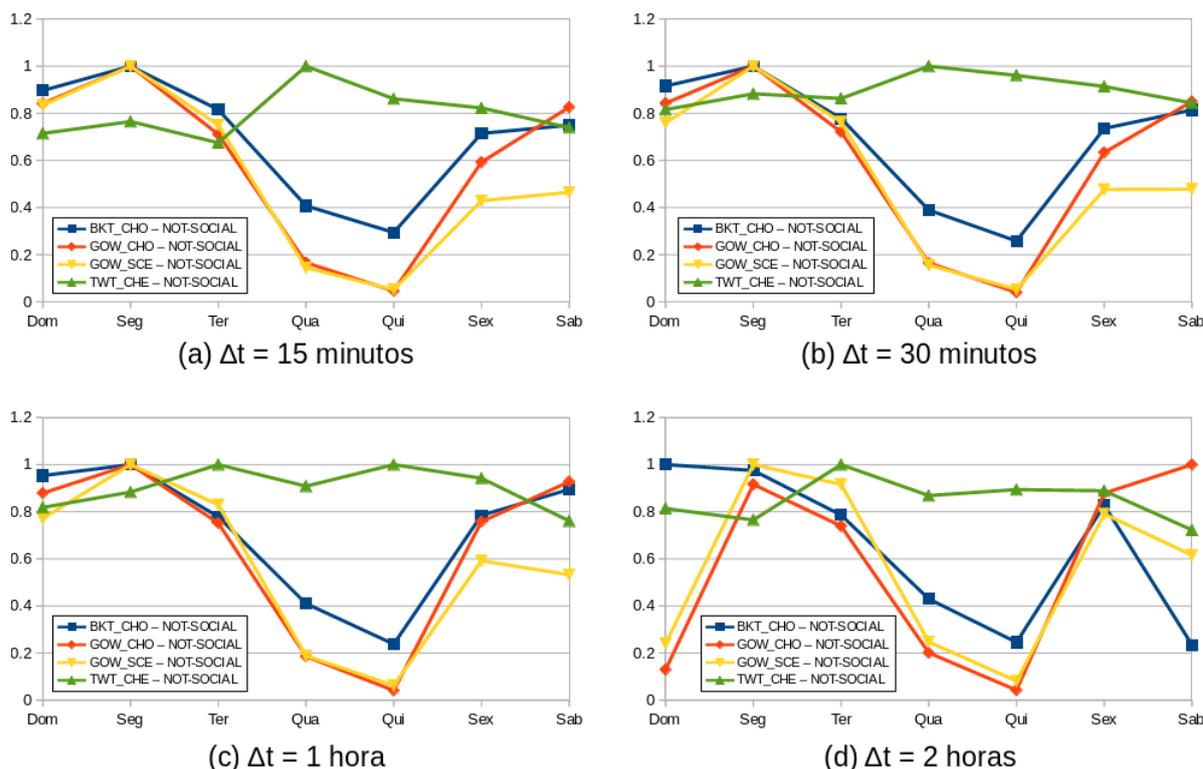
Diante disso, nota-se que as únicas bases em que seria possível a construção de redes SiST por cada dia da semana, são as bases BKT\_CHO, GOW\_CHO, GOW\_SCE e TWT\_CHE. A base TWT\_EIN não foi utilizada devido aos *check-ins* disponibilizados na mesma terem ocorrido em apenas seis dias. A base 4QR\_YUA não foi utilizada porque seria impossível saber qual dia da semana ocorreu um *check-in* uma vez que os referidos dados tiveram as datas anonimizadas pelos autores. No que diz respeito à construção de redes SiST sociais (i.e. representando encontros sociais), só foi possível realizar a análise para as bases BKT\_CHO, GOW\_CHO e GOW\_SCE, uma vez que essas foram as únicas a disponibilizar os laços sociais dos usuários.

A descrição quantitativa do número de nós e conexões, do grau de cada rede criada, do diâmetro da rede, do coeficiente de agrupamento e da média do menor caminho é apresentada no Apêndice B. Os valores dessas propriedades foram calculados por meio do pacote *igraph*<sup>6</sup> de R<sup>7</sup>.

Primeiro, pode-se observar que as redes SiST não sociais dos finais de semana (i.e. sábado e domingo) apresentam grau médio com um valor maior do que aquelas redes geradas a partir de encontros de dias de semana (i.e. de segunda a sexta). A Figura 5.13 apresenta quatro gráficos, cada qual para um intervalo de encontro diferente. Estes gráficos relacionam o grau médio das redes criadas para cada base e para cada dia da semana. Os valores apresentados nos referidos gráficos foram normalizados pelo valor máximo para efeito de comparação.

<sup>6</sup><http://igraph.org/r/>

<sup>7</sup><https://www.r-project.org/>

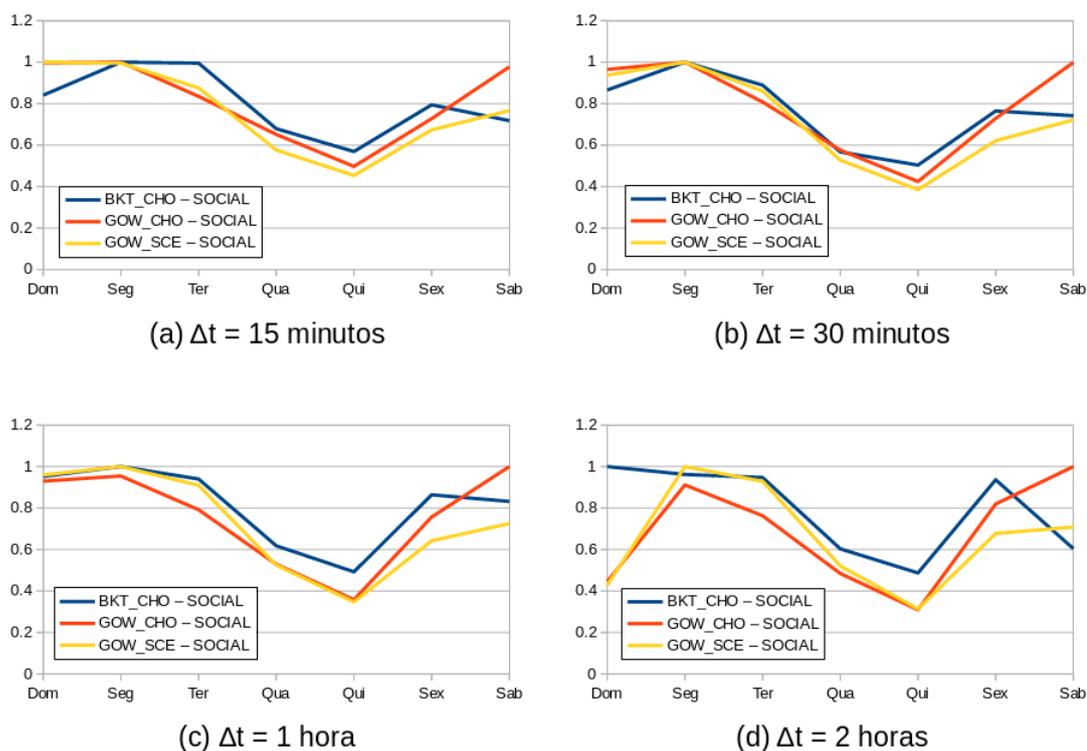


**Figura 5.13:** Distribuição do grau médio das redes SiST não sociais por dia de semana.

No geral, o grau médio das redes cresce entre a sexta e a segunda-feira, indicando que entre esses dias há uma incidência maior de conexão entre os nós, i.e. maior existência de encontros. Este resultado é semelhante ao da análise temporal de encontros (Seção 4.5.3). Nota-se também que as redes SiST construídas a partir de dados de redes sociais originalmente baseadas em localização, i.e. BKT\_CHO, GOW\_CHO e GOW\_SCE, apresentam um padrão comum na distribuição do grau médio para cada dia da semana, diferentemente das redes criadas a partir de TWT\_CHE. Essa diferença na distribuição do grau médio pode ser explicada pelo fato de que os dados de TWT\_CHE não expressam todos os *check-ins*, uma vez que nem todos os usuários de redes sociais baseadas em localização definem a opção para que os *check-ins* sejam republicados no Twitter, conforme já demonstrado na Seção 4.4.

No que diz respeito à distribuição do grau médio de redes SiST sociais, pode-se notar que embora exista uma incidência regular de encontros sociais durante a semana (mais do que a incidência de encontros não sociais, inclusive), existe, também, uma clara tendência de que os encontros sociais sejam postados em finais de semana. Isto indica que apesar de as pessoas publicarem com mais frequência seus encontros sociais durante os fins de semana, este tipo de encontro também é publicado com regularidade durante dias de semana. A Figura 5.14 apresenta a distribuição semanal do grau médio das redes SiST sociais.

A Figura 5.14 mostra que quanto maior o intervalo de encontro, maior é a curva representativa do grau das redes. Ou seja, pode-se perceber um aumento do intervalo de encontro que resultou em um aumento, também, no desvio padrão da média de encontros. Esse aumento

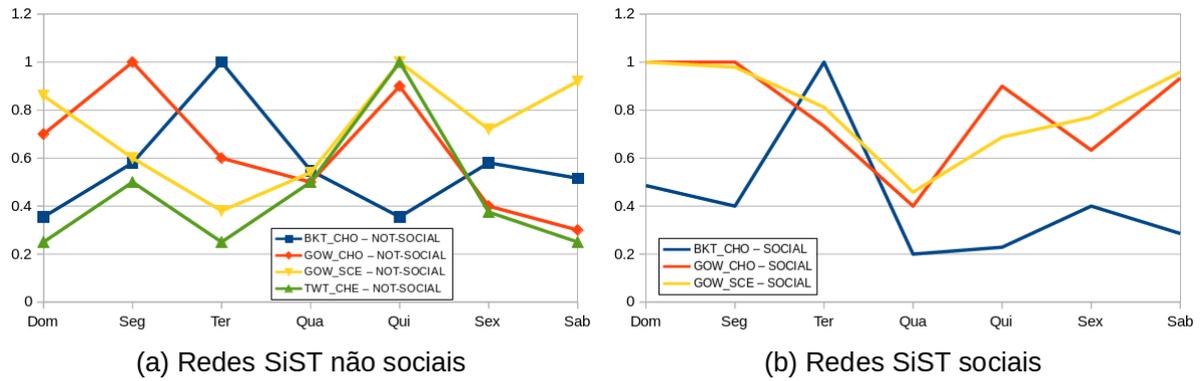


**Figura 5.14:** Distribuição do grau médio das redes SiST sociais por dia de semana.

trouxe uma maior diferença no número de encontros para alguns dias em específico, como por exemplo para o domingo. No sábado, entretanto, nota-se que quando  $\Delta t = 2$  horas, a proporção de encontros cai drasticamente quando comparado aos outros dias. Ao se comparar esse resultado com o da Seção 4.5.5, em que a duração dos encontros foi investigada, pode-se concluir que isso ocorre porque os encontros sociais, em sua maioria, são preferencialmente publicados em até uma hora de intervalo de encontro. Assim, pode-se pressupor que quando o intervalo de encontro é maior do que uma hora, sua publicação não foi feita de forma proposital, ou seja, no mundo físico, os usuários não decidiram publicar, de fato, o encontro.

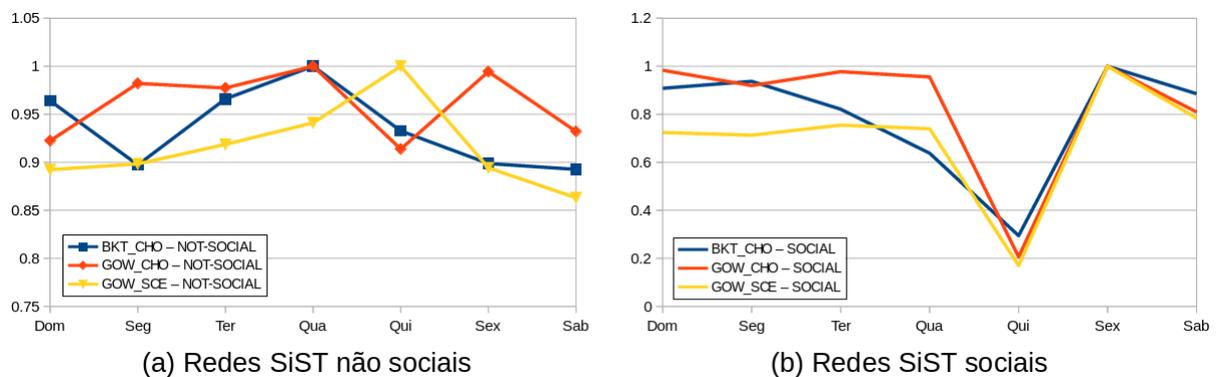
Não foi observada qualquer correlação da ocorrência de encontros com o diâmetro de redes SiST. A Figura 5.15(a) apresenta a distribuição desta propriedade para redes SiST não sociais e a Figura 5.15(b)) para as redes SiST sociais. Em ambas situações as redes foram construídas considerando-se um  $\Delta t = 15$  minutos. Sabe-se que o diâmetro de um grafo representa a maior distância entre dois nós. Desta forma, presume-se que esta falta de correlação esteja relacionada ao alto grau de sensibilidade que essa propriedade tem em redes SiST, uma vez que foi notado que essas redes são caracterizadas por serem grafos não conexos e a consequência da baixa conectividade, por sua vez, implica diretamente nas propriedades relacionadas à distância entre nós, especialmente o diâmetro, que envolve a maior distância entre dois nós. Essa baixa conectividade foi demonstrada na rede da Figura 5.9, em que existe um alto número de nós que não estão conectados com as componentes maiores da rede SiST como um todo.

Como esperado, as redes SiST não sociais dos finais de semana apresentaram uma média



**Figura 5.15:** Diâmetro de redes SiST para  $\Delta t = 15$  minutos.

de menor caminho entre os nós menor do que aquelas dos outros dias da semana. Essa tendência correlaciona-se com o fato de que quanto maior a ocorrência de encontros, mais nós e conexões as redes passam a conter e, conseqüentemente, menor a distância entre os nós. Apesar de essa tendência ter sido apresentada para as redes geradas a partir de todos os intervalos de encontro, a Figura 5.16(a) apresenta um gráfico que mostra apenas as redes SiST não sociais construídas a partir do intervalo de encontro igual a 1 hora. Embora outros gráficos semelhantes possam ser construídos a partir dos dados do Apêndice B, estes gráficos não são apresentados aqui por serem semelhantes ao gráfico da referida figura.



**Figura 5.16:** Distância de redes SiST para  $\Delta t = 1$  hora.

Diferentemente de como ocorreu com as redes SiST não sociais, não foi possível observar qualquer tendência no que diz respeito à média do menor caminho das redes SiST sociais. A Figura 5.16(b) apresenta essa medida. A explicação para essa divergência pode ser, novamente, a baixa conectividade das redes SiST, uma vez que as redes SiST sociais são menores em tamanho e em densidade quando comparadas com as redes SiST não sociais.

Finalmente, a análise do coeficiente de agrupamento das redes permitiu observar, primeiramente, que as redes SiST apresentam um alto valor médio para esta medida. Foi possível chegar a esta conclusão quando se comparou cada rede SiST com redes aleatórias de mesmo tamanho, i.e. redes com mesmo número de nós e conexões. Em uma rede com alto coeficiente

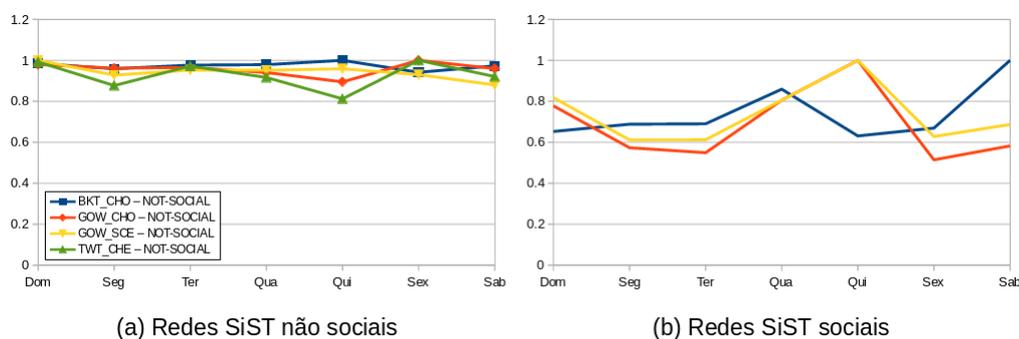
de agrupamento, seus usuários tendem a se agrupar. Ou seja, usuários conectados a um usuário com alto coeficiente de agrupamento tendem a possuir conexões entre si. Em outras palavras: usuários que se encontram com um usuário com um alto coeficiente de agrupamento tendem, também, a se encontrarem.

A Tabela 5.6 apresenta as propriedades de coeficiente de agrupamento, a média do menor caminho e o diâmetro de redes aleatórias de mesmo tamanho das redes SiST não sociais de BKT\_CHO construídas a partir de um intervalo de encontro igual a 15 minutos. As redes originais, i.e. aquelas de BKT\_CHO, têm suas propriedades apresentadas no Apêndice B, especificamente na Tabela B.1. As tabelas com propriedades de outras redes aleatórias de mesmo tamanho das redes SiST originais são apresentadas no Apêndice C.

Rede	coef. agrup.	distância	diâmetro
Dom	0.01143474	6.143574	14
Seg	0.01723821	5.711992	14
Ter	0.02487547	6.46397	15
Qua	0.008197279	7.373392	21
Qui	0.01923077	3.045783	10
Sex	0.009982684	8.482614	24
Sab	0.01354583	7.803067	19

**Tabela 5.6:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.1.

Pode-se perceber que para as redes SiST sociais, a tendência de os usuários formarem grupos entre si é ainda maior nos finais de semana. Essa observação é resultante do fato de que esse tipo de rede possui tendência de crescimento do coeficiente de agrupamento médio nas redes SiST sociais construídas com encontros ocorridos em finais de semana. A Figura 5.17 apresenta os gráficos comparando a média do coeficiente de agrupamento de cada rede SiST construída com um intervalo de encontro igual a 2 horas.



**Figura 5.17:** Coeficiente de agrupamento das redes SiST.

### 5.6.1.2 Investigação sobre leis que governam as conexões das redes SiST

A investigação sobre leis que governam as conexões das redes SiST se iniciou pela análise da distribuição do grau dessas redes. Essa análise foi feita considerando as redes SiST

como sendo redes não direcionais e utilizando-se da mesma abordagem com que foi feita a análise da distribuição de *check-ins* e encontros, tal qual descrito na Seção 4.5.1 e Seção 4.5.3, respectivamente. Ou seja, foi executado um processo de inferência estatística sobre a distribuição do grau médio de cada uma das redes com o objetivo de discernir se o modelo *power-law* se ajusta bem ao conjunto de dados ou se outro modelo, como o log-normal, se ajusta melhor.

A investigação apresentada nesta seção fez uso das mesmas redes da análise descrita na Seção 5.6.1.1, além de redes adicionais providas das bases 4QR\_YUA e TWT\_EIN. No entanto, para que os métodos de inferência estatística pudessem ser executados de forma correta, foram utilizadas apenas redes em que o número de conexões fosse maior do que 100. Neste caso, para estas últimas bases de dados, apenas se enquadram nesta prerrogativa as redes SiST construídas considerando-se um intervalo de encontro igual a 2 horas.

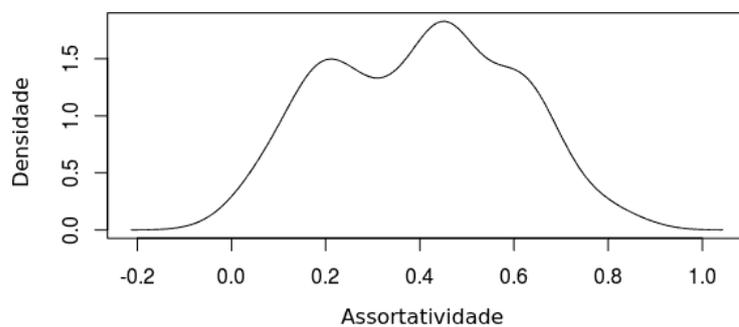
No teste de inferência aplicado sobre as redes SiST, foi escolhido um nível de significância igual a 90%. Primeiro, o teste de inferência foi realizado com o objetivo de ajustar o modelo *power-law* à distribuição do grau médio das redes e, depois disso, de investigar se outro modelo teórico, que não *power-law*, se ajusta melhor ao referido conjunto de dados. Para este último processo de inferência, foi utilizado, novamente, o Teste de Vuong.

Os dados quantitativos do teste realizado para investigação aqui descrita estão inteiramente apresentados no Apêndice D. Ao se analisar a função acumulada do grau dos nós de cada uma das redes descritas no referido apêndice, bem como os resultados do teste de inferência, pode-se chegar à conclusão de que o modelo que melhor se ajusta à distribuição do grau médio de redes SiST é o modelo *power-law*, uma vez que essa distribuição se ajusta bem a cerca de 60% das redes SiST analisadas. Há casos em que as redes não foram aproximadas por uma *power-law* ou foram aproximadas por uma *log-normal* (o que representa cerca de 9% de todas as redes) ou o resultado da análise foi inconclusivo (aproximadamente 31% dos resultados).

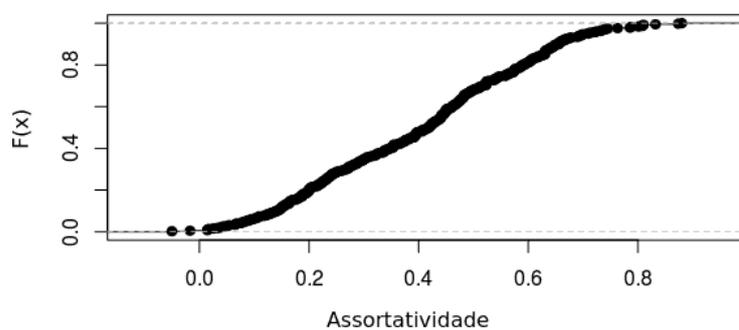
O fato de as redes SiST serem governadas por uma *power-law* indica que usuários de redes sociais online não publicam seus encontros aleatoriamente, ou seja, existe um processo de criação de conexão preferencial baseado em alguma característica, seja dos usuários envolvidos no encontro, seja do local em que ocorreu o encontro, por exemplo: a influência de algum usuário participante do encontro ou a popularidade do local escolhido para publicar o encontro. Esta característica seria semelhante se outro modelo de “cauda pesada” (i.e. *heavy-tailed*) se ajustasse bem ao conjunto de dados, como é o caso da log-normal, por exemplo, embora seja mais proeminente em redes governadas por uma *power-law*.

Com o objetivo de se compreender ainda mais as leis que governam encontros, se investigou o coeficiente de assortatividade das redes. Esse coeficiente é listado nas tabelas do Apêndice B. A Figura 5.18(a) mostra a função de densidade desta propriedade e a Figura 5.18(b) ilustra a função de distribuição acumulada desta mesma medida. Pode-se observar que a assortatividade é positiva, sendo a maioria dessa medida (aproximadamente 60%) concentrada entre 0 e 0.5. Este resultado indica que as redes SiST são assortativas, i.e. um nó de grau elevado preferencialmente se conecta com nós de graus elevados e, do mesmo modo, nós de grau baixo

tendem a se conectar com nós de grau baixo, embora este último ocorra com menor regularidade do que na primeira situação, o que é uma característica da *power-law*.



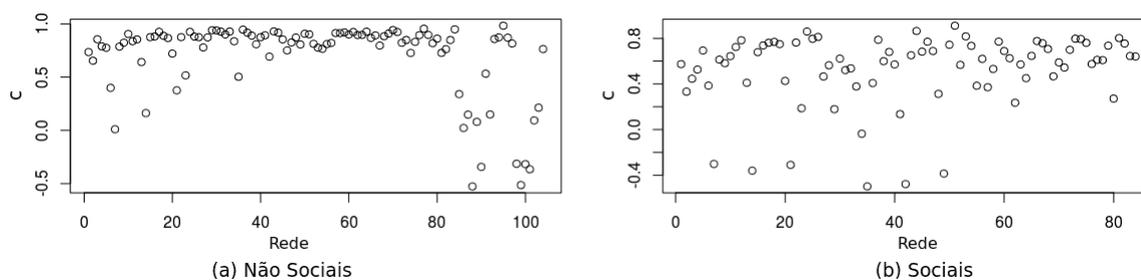
(a) Função de Densidade



(b) Função de Distribuição Acumulada

**Figura 5.18:** Análise da assortatividade das redes SiST.

Como a assortatividade é calculada por meio da correlação do grau de entrada e saída de cada nó, essa correlação foi calculada utilizando-se o método de correlação implementado em R. Neste caso, pode-se notar que o valor da correlação se concentra, de fato, entre 0.5 e 1.0. A Figura 5.19 plota a distribuição da correlação entre os nós das redes, sendo a Figura 5.19(a) relacionada às redes SiST não sociais e a Figura 5.19(b) às redes SiST sociais. Este resultado em conjunto com a análise da assortatividade demonstra que, realmente, os usuários preferem se conectar (i.e. publicar seus encontros) com outros usuários considerados “iguais”, ou seja, existe uma correlação positiva entre o grau de entrada e o grau de saída dos nós das redes SiST.



**Figura 5.19:** Correlação in-degree e out-degree de redes SiST.

Finalmente, as redes SiST foram investigadas quanto à possibilidade de apresentarem

o fenômeno do mundo pequeno (*small world networks*). Redes com esse fenômeno possuem duas características principais: primeiro, essas redes possuem um coeficiente de agrupamento muito alto, segundo, essas redes apresentam um valor pequeno para a média do menor caminho entre os nós. Além disso, uma rede do mundo pequeno é caracterizada também pelo fato de suas conexões serem criadas de acordo com um processo governado por uma função *power-law*.

Até aqui, sabe-se que além do alto coeficiente de agrupamento, redes SiST podem ser modeladas por uma *power-law*. Para se interpretar se a média do menor caminho de redes SiST possui, de fato, um valor pequeno, se comparou essas redes com redes aleatórias como visto na Seção 5.6.1.1. Diante dessa investigação, tal qual sugerido por (WATTS; STROGATZ, 1998), pode-se concluir que redes SiST são redes do mundo pequeno.

O fato de redes SiST serem consideradas redes do mundo pequeno sugere que entre dois usuários quaisquer, existe uma média muito pequena de usuários que se conhecem e, conseqüentemente, publicam encontros. De acordo com o fenômeno do mundo pequeno, tal qual demonstrado na Seção 2.2.3.3, essa média pode ser de seis nós, ou seja, seis usuários. Ou seja, apesar de uma fração muito pequena de usuários possuírem encontros em comum com outros usuários, essa quantidade de encontros é muito alta para essa fração de usuários, o que acaba por aumentar a probabilidade de dois usuários quaisquer possuírem encontros em comum entre si.

### 5.6.2 Detecção de Padrões de Mobilidade de Usuários

Para realização dos experimentos relativos à mobilidade de usuários, no lugar de redes SiST diárias, foram construídas redes por meio dos *check-ins* ocorridos durante os 15 dias em que houve maior representatividade de *check-ins* de cada base. Isto foi feito para tentar identificar padrões que ocorressem durante um intervalo maior do que, simplesmente, um dia.

Os algoritmos de detecção de padrões de grupo de encontros e de liderança de encontros detectam instâncias dos respectivos padrões utilizando-se a matriz de adjacência das redes SiST. Entretanto, antes de se investigar por esses padrões, os encontros duplicados foram detectados, já que esses encontros podem compor as instâncias dos padrões de mobilidade aqui analisados. Diante da impossibilidade de se descrever todos os padrões detectados, optou-se por apresentar apenas um exemplo de cada um dos padrões.

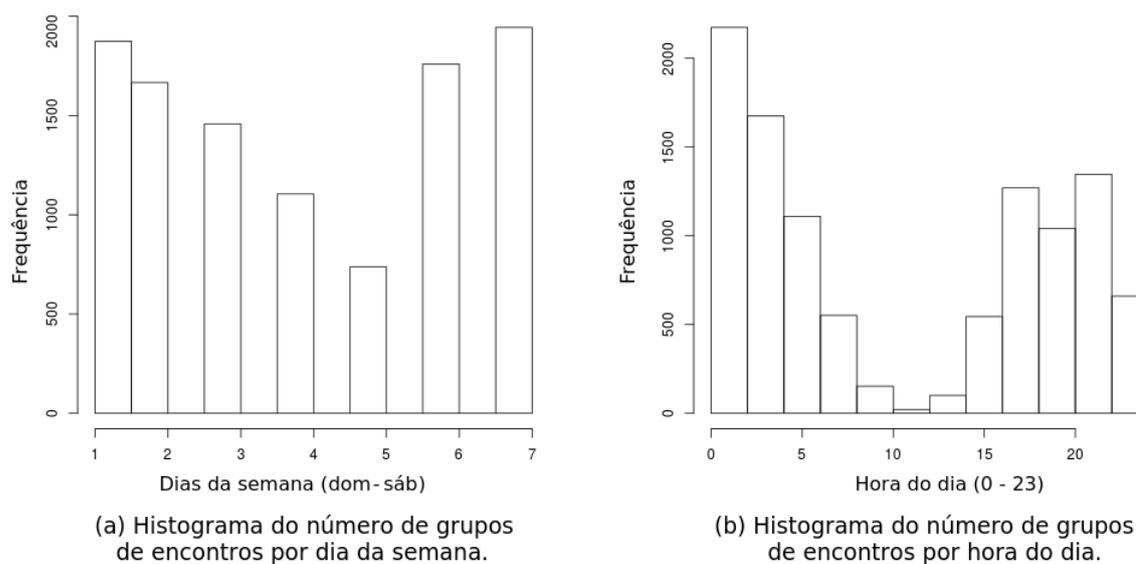
O exemplo de encontros duplicados foi detectado em redes SiST sociais com o intervalo de encontro igual a 15 minutos para a base BKT\_CHO. A Tabela 5.7 apresenta uma porção dos *check-ins* em que foram detectados encontros duplicados. Os *check-ins* envolvidos no padrão citado estão destacados em negrito. Nota-se que neste caso os *check-ins* dos usuários 13 e 19 ocorridos no local 260405 (i.e. os *check-ins* de id igual a 3388651, 3388650, 3388815 e 3388814) podem ser caracterizados como uma instância do padrão de encontros duplicados, uma vez que os mesmos foram publicados de acordo com a Definição 6, i.e. de que a diferença temporal entre os *check-ins* repetidos do precursor é menor ou igual ao intervalo de encontro.

Os grupos de encontros também foram detectados. Pode-se notar que independentemente

$id_{ck}$	$id_{user}$	$id_{local}$	instante de tempo
2996813	13	200387	2009-03-18 04:42:28
2996903	19	203113	2009-03-18 04:43:32
<b>3388651</b>	<b>13</b>	<b>260405</b>	<b>2009-03-18 05:19:47</b>
<b>3388650</b>	<b>13</b>	<b>260405</b>	<b>2009-03-18 05:20:41</b>
<b>3388815</b>	<b>19</b>	<b>260405</b>	<b>2009-03-18 05:21:54</b>
<b>3388814</b>	<b>19</b>	<b>260405</b>	<b>2009-03-18 05:23:14</b>
3388648	13	260405	2009-03-18 06:26:33-03
3388647	13	260405	2009-03-18 06:28:29-03
3066127	19	215798	2009-03-18 08:18:54-03

**Tabela 5.7:** Exemplo de padrão de duplicidade de encontros detectado. Check-ins relacionados ao padrão estão em negrito.

de serem sociais ou não, grupos de encontros ocorrem principalmente nos finais de semana e durante a madrugada, especialmente entre meia noite e 2 horas da manhã. A Figura 5.20(a) apresenta o histograma da média de grupos de encontros por dias de semana e a Figura 5.20(b) ilustra o histograma da média de encontros distribuídos por hora do dia.



**Figura 5.20:** Histogramas do número de encontros por dia de semana e por hora do dia.

A Tabela 5.8 destaca uma instância de um grupo de encontros detectado em uma das bases de dados. Neste caso, o grupo é composto pelos usuários de identificadores 30, 49, 102 e 105 que se encontraram no local 6624. Nota-se, portanto, que cada usuário do referido grupo publicou seu *check-in* respeitando o intervalo de encontro de 15 minutos, razão pela qual esses *check-ins* foram considerados como participantes do grupo de encontros. Pode-se observar, também, que para que o grupo de encontros pudesse ser detectado, os encontros duplicados foram capturados previamente. Este é o caso, por exemplo, dos *check-ins* de *id* igual a 4037226, 3843874 e 4037225, em que ocorreram no mesmo local e, devido à repetição do *check-in* do usuário precursor (i.e. usuário de *id* = 30), foram caracterizados como encontros duplicados.

$id_{ck}$	$id_{user}$	$id_{local}$	instante de tempo
355563	102	126583	2009-03-16 01:58:19
404088	30	126583	2009-03-16 02:00:45
4157101	49	26214	2009-03-16 04:31:40
<b>4037226</b>	<b>30</b>	<b>6624</b>	<b>2009-03-16 05:51:26</b>
<b>3843874</b>	<b>102</b>	<b>6624</b>	<b>2009-03-16 05:54:51</b>
<b>4037225</b>	<b>30</b>	<b>6624</b>	<b>2009-03-16 05:57:18</b>
<b>4037695</b>	<b>49</b>	<b>6624</b>	<b>2009-03-16 06:01:51</b>
<b>3843942</b>	<b>105</b>	<b>6624</b>	<b>2009-03-16 06:03:09</b>
577082	102	168202	2009-03-16 07:17:57

**Tabela 5.8:** Exemplo de padrão de grupo de encontros detectado. Check-ins relacionados ao padrão estão em negrito.

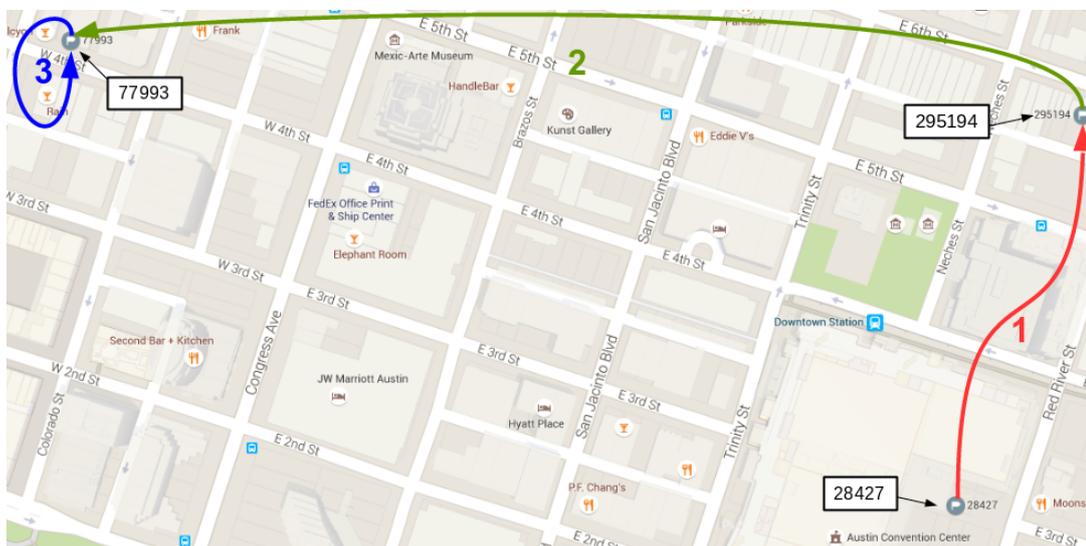
Finalmente, a Tabela 5.9 apresenta um conjunto de *check-ins* dentre os quais estão presentes alguns que foram caracterizados como instâncias de um padrão de liderança de encontros. Estes *check-ins* estão destacados em negrito. Pode-se observar, portanto, que o usuário de identificador 36 claramente publica *check-ins* imediatamente depois (i.e. respeitando-se  $\Delta t$ ) do usuário de identificador igual a 1511. Conforme a definição do referido padrão, nota-se que os *check-ins* em destaque na Tabela 5.9 foram assim identificados porque seus usuários se encontraram consecutivamente em três locais diferentes.

$id_{ck}$	$id_{user}$	$id_{local}$	instante de tempo
1931389	10378	21959	2009-03-13 02:22:46
3259610	2460	253767	2009-03-13 02:23:02
<b>4104801</b>	<b>1511</b>	<b>28427</b>	<b>2009-03-13 02:23:13</b>
<b>1963109</b>	<b>36</b>	<b>28427</b>	<b>2009-03-13 02:36:21</b>
4061023	7237	14775	2009-03-13 02:37:05
1133010	180	269439	2009-03-13 02:37:39
<b>1323823</b>	<b>1511</b>	<b>295194</b>	<b>2009-03-13 02:47:01</b>
<b>3586872</b>	<b>36</b>	<b>295194</b>	<b>2009-03-13 03:00:31</b>
108453	1523	77759	2009-03-13 03:03:41
1259539	993	287700	2009-03-13 03:03:48
<b>108428</b>	<b>1511</b>	<b>77993</b>	<b>2009-03-13 04:27:00</b>
<b>2290749</b>	<b>36</b>	<b>77993</b>	<b>2009-03-13 04:40:02</b>
994208	21240	237258	2009-03-13 04:42:25
<b>108427</b>	<b>1511</b>	<b>77993</b>	<b>2009-03-13 06:29:47</b>
<b>2290747</b>	<b>36</b>	<b>77993</b>	<b>2009-03-13 06:35:35</b>
3453095	1168	291215	2009-03-13 06:41:39

**Tabela 5.9:** Exemplo de padrão de grupo de liderança de encontros detectado. Check-ins relacionados ao padrão estão em negrito.

Pode-se notar que em um intervalo de aproximadamente quatro horas, quatro encontros foram publicados pelos usuários 1511 e 36, sendo sempre o usuário 1511 o precursor. A instância do padrão de liderança da Tabela 5.9 é ilustrada no mapa da Figura 5.21. Neste mapa, nota-se

que os três locais em que houve a sequência de encontros são próximos, inclusive, do Centro de Convenções de Austin, local em que foi detectado uma grande quantidade de encontros, tal qual descrito no Capítulo 4. As setas do mapa indicam as direções com que os referidos usuários se locomoveram no mundo físico para que os encontros ocorressem conforme foram publicados.



**Figura 5.21:** Mapa apresentando os locais em que o padrão de liderança de encontro descrito foi detectado.

A identificação do padrão de liderança pode auxiliar na interpretação sobre se o usuário considerado líder possui, de fato, poder de influência na publicação em redes sociais com serviços de localização. Para se fazer esse tipo de análise, pode-se utilizar dados contextuais a partir das redes sociais, como informações do perfil dos usuários ou informações sobre os locais em que os encontros ocorreram. Esses padrões podem auxiliar no processo de definição de novos serviços baseados em redes sociais online ou na investigação de padrões de mobilidade urbana. Por exemplo, pode-se estudar se esses padrões são recorrentes em diferentes cidades ou se foram identificados em locais de mesmo tipo e/ou respeitando algum padrão temporal. As instâncias dos padrões de liderança podem ser utilizados para relacionar os diferentes tipos de locais com o poder de influência dos usuários nestes locais. Por exemplo, pode-se investigar se uma pessoa tem alto poder de influência na liderança de encontros em restaurantes, mas pode não ter esse poder de influência quando realiza *check-ins* em cinemas ou shows musicais, por exemplo.

Para se interpretar esses padrões, no entanto, deve-se notar que publicação de encontros em redes sociais não necessariamente implica que as pessoas estão de fato juntas, se relacionando no mundo físico, uma vez que dependendo do local ou do tipo do local elas sequer podem ter se encontrado. Entretanto, publicar que estão no mesmo local pode trazer diversas consequências em estudos relacionados a análise de redes sociais, como a indicação de que os usuários possuem um alto grau de semelhança e a consequente sugestão de locais e/ou amigos baseado no histórico de visitas dos mesmos, por exemplo. Além disso, deve-se observar também que a ordem temporal da publicação não representa, de fato, a ordem com que os usuários chegaram ao local físico

em que estão visitando. Diante disso, os padrões devem ser analisados e interpretados à luz dos possíveis impactos na influência dos usuários nas mídias sociais e não no mundo físico.

### 5.7 Considerações Finais

Este capítulo definiu o modelo SiST, seus conceitos básicos e os algoritmos implementados tanto para que as redes SiST fossem instanciadas bem como para que os padrões de mobilidade fossem detectados. O modelo foi validado por meio de diferentes abordagens: *(i)* foram construídas redes a partir do modelo proposto; *(ii)* essas redes foram analisadas estatisticamente e comparadas por meio da análise apresentada no Capítulo 4 e, finalmente, *(iii)* as redes instanciadas foram estudadas neste capítulo com o objetivo de se detectar padrões de encontro e, conseqüentemente, padrões de mobilidade de usuários. Esse processo de validação do modelo gerou resultados significativos e que demonstraram a aplicabilidade do mesmo em tarefas de análise de redes sociais online.

O modelo, entretanto, pode oferecer resultados ainda mais significativos se aplicado considerando-se informações contextuais adquiridas a partir das redes sociais. Essas informações não foram utilizadas no processo de validação devido à impossibilidade de obtê-las a partir das bases de dados utilizadas. Por exemplo, pode-se utilizar as informações obtidas a partir das redes SiST e dos algoritmos aqui propostos em conjunto com informações sobre o histórico do uso das redes sociais dos usuários bem como o histórico sobre quais usuários têm frequentado determinados locais para se extrair conhecimento a respeito da mobilidade dos usuários bem como a respeito do perfil dos usuários em relação aos locais visitados. Com esses resultados, os analistas de redes sociais podem estar aptos a definir estratégias para alcançar potenciais consumidores de produtos ou, simplesmente, para auxiliar no processo de divulgação de determinado estabelecimento por meio de pessoas com alto poder de influencia em redes sociais providas com serviços de localização.

## 6

### Conclusões

O imenso repositório de dados que as redes sociais online têm criado, tem feito com que diversos estudos surjam com o objetivo de analisar e estudar esses dados sob as mais variadas perspectivas. A tese apresentada neste documento propõe investigar esse vasto repositório sob um ponto de vista diferente do que se tem visto na literatura: o ponto de vista de encontros publicados por usuários.

Os trabalhos relacionados demonstraram que os fenômenos espaço-temporais se sobrepõem aos fenômenos sociais quando se deseja estudar, por exemplo, o modo como uma informação flui em uma rede social. Nesse sentido, portanto, a análise de encontros publicados por usuários pode ser de grande relevância para a área de análise de redes sociais como um todo uma vez que encontros são capazes de englobar tanto as informações sociais como as espaço-temporais. Isto é reforçado quando se analisa *check-ins* e encontros separadamente e percebe-se que esses tipos de dados são caracterizados por padrões de publicação diferentes e, inclusive, por leis estatísticas distintas. Diante de todas as definições e resultados descritos nessa tese, esse capítulo tem o objetivo de apresentar as conclusões do referido trabalho. Este capítulo está organizado como segue. A Seção 6.1 descreve as principais contribuições e a Seção 6.2 lista os trabalhos futuros.

#### 6.1 Principais Contribuições

Para se estudar os padrões relacionados à publicação de encontros e se compreender como os mesmos, de fato, ocorrem, foram analisados primeiramente os *check-ins* independentemente da detecção desses encontros. Nesse estudo foi possível observar que a publicação de *check-ins* é modelada por meio de uma log-normal. Esse resultado condiz com os resultados de alguns trabalhos, conforme demonstrado no Capítulo 3, quando mostraram que as relações sociais são modeladas por uma *power-law* e as interações espaciais por uma log-normal.

O fato de *check-ins* serem governados por uma log-normal e não por uma *power-law* traz diversas consequências à interpretação desse fenômeno das redes sociais online. Primeiro, como log-normal é uma função de cauda pesada, é notório que existem poucos usuários que publicam uma grande quantidade de *check-ins* e, por outro lado, existem muitos usuários que

publicam poucos *check-ins*. Entretanto, essa diferença não é tão latente como ocorre com as interações sociais, que normalmente são governadas por meio de uma *power-law*. Ou seja, os *check-ins* são dotados de certa regularidade quanto a sua publicação. Uma explicação para isso é o fato do *check-in* ser uma funcionalidade que envolve interações espaciais, ou seja, envolve deslocamentos físicos dos indivíduos, diferentemente das interações sociais em redes online. Além do estudo sobre a probabilidade dos *check-ins*, foram investigados padrões temporais de publicação. Nesse sentido, foi observado que os *check-ins* são publicados, normalmente, em meses de alta temporada (i.e. meses de férias, como no meio e no final do ano), durante finais de semana e, particularmente, entre as 22 horas e 2 horas da madrugada.

Tomando-se conhecimento a respeito das publicações de *check-ins*, o algoritmo de detecção de encontros foi executado com o objetivo de investigar os encontros para cada uma das bases de dados estudadas. Esses encontros foram detectados por meio de diferentes intervalos de encontros e considerando-se ou não os laços sociais quando esses estavam disponíveis. Diante da execução do algoritmo responsável pela detecção de encontros, foi possível realizar, também, um estudo de probabilidade desses encontros.

Esse estudo de probabilidade trouxe algumas conclusões sobre a ocorrência de encontros. Primeiro, o que se pode notar é que assim como *check-ins*, encontros não ocorrem aleatoriamente. Tanto a análise da distribuição de encontros por pares de usuários como por locais permitiu observar que a publicação de encontros é governada por uma *power-law*. Essa conclusão reforça o que já foi descrito nos trabalhos relacionados, quando foi demonstrado que uma função *power-law* pode ser uma boa escolha para se modelar relações sociais entre usuários. Diante disso, como a dimensão social é inerente à publicação de encontros, esse evento pode ser, portanto, modelado por meio de uma *power-law*.

A análise da distribuição temporal de encontros permitiu observar que esse tipo de evento ocorre, normalmente, em fins de semana e, especialmente, durante a madrugada, entre as 23h e 3 horas da manhã. Foi nessa análise temporal, inclusive, que foi possível detectar um evento que durante a análise de *check-ins* não foi detectado: o SXSW na cidade de Austin. Finalmente, a análise da duração de encontros permitiu observar que cerca de 70% dos encontros sociais são publicados em um intervalo de tempo entre os *check-ins* menor ou igual a uma hora.

Além do estudo de probabilidade, esta tese propôs também um modelo de representação de encontros baseado na teoria dos grafos: o modelo SiST. Diferentemente dos outros modelos já existentes na área de redes sociais online, o modelo SiST permite que as interações espaço-temporais entre usuários sejam representadas e analisadas de acordo com o ponto de vista de encontros entre usuários. O modelo é caracterizado pela sua simplicidade e por prover definições e formalizações nos diferentes níveis de abstração: seja ele conceitual, lógico ou físico.

O modelo SiST foi apresentado no Capítulo 5 e foi avaliado em termos de redes construídas a partir do mesmo conjunto de dados utilizados na análise da distribuição de *check-ins* e de encontros, descrita no Capítulo 4. Essas redes foram validadas, inicialmente, por meio de diferentes pontos de vistas com o objetivo de comparar seus resultados com os da análise

anterior, isto é, do Capítulo 4.

A análise das propriedades das redes SiST apresentada no Capítulo 5 demonstrou resultados equivalentes àqueles obtidos a partir da análise da distribuição de encontros do Capítulo 4. Por exemplo, redes SiST construídas com dados de finais de semana apresentaram maior coeficiente de agrupamento do que aquelas com dados de outros dias da semana. Isso indica, claramente, a tendência de usuários de publicarem seus encontros nos finais de semana, assim como demonstrado na análise do grau médio das redes SiST e, também, na análise da distribuição temporal de encontros.

Na investigação sobre as leis que, possivelmente, governam as conexões das redes SiST, pode-se notar que publicação de encontros, de fato, é governada por uma *power-law*, confirmando também os resultados do estudo apresentado na análise da distribuição de encontros. Esse fato indica, novamente, que encontros não são publicados aleatoriamente.

A análise da assortatividade e os resultados provenientes da comparação das redes SiST com redes aleatórias, permitiu concluir que redes SiST são redes caracterizadas pelo fenômeno do mundo pequeno. Ou seja, nós das redes SiST se conectam uns aos outros respeitando um mecanismo de conexão preferencial. A análise da assortatividade das redes SiST, inclusive, demonstrou que os valores dessa medida são equivalentes aqueles demonstrados em outros estudos na área de redes sociais, conforme pode ser visto no Capítulo 3. O fato de redes SiST serem do mundo pequeno, por sua vez, indica que elas exibem características que são ideais para colaboração e fluxo de informação, uma vez que redes do mundo pequeno são redes que são caracterizadas pelos seis graus de separação.

Finalmente, as redes SiST também foram avaliadas no que diz respeito à detecção de padrões de mobilidade de usuários. Nessa avaliação, foram definidos três padrões onde um deles, o padrão de encontros duplicados, diz respeito a um padrão de publicação de encontros em que usuários republicam encontros com outros usuários. Além do padrão de encontros duplicados, os padrões de grupos de encontros e de liderança de encontros também foram definidos. Além das definições desses padrões, foram propostos algoritmos capazes de detectá-los considerando uma rede SiST como argumento desses algoritmos.

Os padrões de mobilidade de usuários propostos nessa tese podem auxiliar no processo de análise de redes sociais de acordo com padrão de uso de usuários deste tipo de ferramenta. Os resultados da detecção desses padrões podem ser aplicados na definição de novos serviços baseados em localização ou em estratégias no planejamento de estratégias de marketing em redes sociais online. Pode-se fazer uso do algoritmo de detecção de grupos de encontros, por exemplo, para detectar eventos ou para estudar padrões temporais de publicação de encontros em grupo. Pode-se, ainda, fazer uso das instâncias do padrão de liderança para se investigar se os usuários considerados “líderes” são, de fato, líderes de encontros no mundo físico considerando-se a disponibilidade de outras informações durante o processo de análise da rede social.

Além da problemática em se construir uma base de dados própria, pode-se considerar que a maior dificuldade desse trabalho foi fazer uso dos dados disponíveis e correlacioná-los com

fontes externas, como por exemplo relacionar dados de encontros com dados sobre o perfil dos usuários ou dos locais em que os encontros ocorreram. Diante de tudo o que foi descrito nessa tese, pode-se responder à pergunta da Seção 1.2 (i.e. “*A publicação de encontros entre usuários de redes sociais online respeita algum padrão de publicação e/ou padrões de mobilidade de usuários?*”) dizendo-se que, sim, as publicações de encontros entre usuários de redes sociais online respeitam padrões de publicações assim como padrões de mobilidade de usuários. A busca por respostas a essa pergunta está fundamentada no que se conseguiu responder a partir das perguntas mais específicas apresentadas naquela seção, como se segue:

P1: De que forma a publicação de *check-ins* se diferencia da publicação de encontros?

A publicação de encontros e *check-ins* se diferencia em diversos aspectos, desde a demonstração dos padrões temporais de publicação assim como pelo fato de ambos eventos serem governados por leis estatísticas diferentes. Os resultados que fundamentam essa resposta estão demonstrados tanto na análise de *check-ins* e encontros do Capítulo 4, bem como reforçados na análise de encontros representados pelas redes SiST do Capítulo 5;

P2: Existe alguma lei estatística que modela a publicação de encontros?

Sim. Os resultados apresentados no Capítulo 4 demonstraram ostensivamente que *check-ins* são governados por uma log-normal e que encontros são governados por uma *power-law*. Os dados provenientes da inferência estatística da lei que rege a distribuição do grau das redes SiST apresentados no Capítulo 5 reforçaram também este último resultado;

P3: Existem fatores temporais que influenciam na publicação de encontros?

Sim. Além de serem publicados com mais frequência em meses de férias e durante finais de semana, os encontros são publicados normalmente entre dez da noite e duas da manhã, seguindo a mesma tendência da publicação dos *check-ins*. Além disso, pode-se observar também que os encontros sociais são caracterizados por serem publicados em um intervalo de tempo pequeno entre *check-ins* envolvendo o encontro;

P4: Existem padrões de movimentação de usuários que estão, possivelmente, relacionados à publicação de encontros dos mesmos?

Para se responder a essa pergunta se faz necessário que sejam feitos outros estudos. Entretanto, existem sinais de que usuários podem ser, sim, influenciados por outros usuários no que diz respeito à publicação de encontros e, conseqüentemente, à movimentação dos mesmos no globo. Pode-se chegar a essa conclusão não somente diante dos padrões temporais de publicação mas também diante da existência de padrões de grupo de encontros e de liderança de encontros;

P5: Existem pessoas que possuem um maior poder de influência do que outras no que diz respeito à publicação de encontros?

Assim como diante da resposta da pergunta P4, para se responder essa pergunta pode-se

utilizar informações provenientes de outras fontes. Entretanto, diante do estudo apresentado nesta tese, pode-se dizer que existem sinais de que certos usuários são, de fato, líderes na publicação de encontros e que usuários seguem uns aos outros no que diz respeito aos locais de publicação daqueles que, na teoria, possuem maior poder de influência. Pode-se chegar a essas diante do grande número de reincidência de usuários considerados “precursores” bem como diante da identificação de usuários “líderes” durante a detecção de instâncias do padrão de liderança de encontros;

P6: Existem locais que possuem um maior poder de agregação de encontros de usuários?

Diante do que foi estudado, aparentemente sim: alguns locais possuem um maior poder de agregação tendo em vista que os mesmos possuem um grande número de *check-ins* associados e também um grande número de encontros. Para se chegar, de fato, à conclusão sobre essa questão, se faz necessário estudos mais aprofundados considerando-se, inclusive, informações de contexto relativas aos locais, como por exemplo o tipo do local ou o histórico daquele local.

Finalmente, diante de tudo o que foi exposto, pode-se resumir as contribuições dessa tese como se segue:

- Análise da probabilidade de um usuário realizar *check-in*;
- Descrição de padrões temporais de publicação de *check-ins*;
- Definição de um algoritmo para detecção de encontros em redes sociais online;
- Análise da probabilidade de dois usuários publicarem um encontro;
- Descrição de padrões temporais de publicação de encontros;
- Definição de um modelo para representar encontros entre usuários de redes sociais online;
- Descrição de padrões e leis que governam publicação de encontros entre usuários;
- Definição de algoritmos para detecção de padrões de mobilidade de usuários;

Essas contribuições podem ser utilizadas tanto no mercado de publicidade e marketing por analistas de redes sociais, como também para estudos referentes aos fenômenos sociais, relacionados à mobilidade de usuários ou aos comportamentos dos usuários. Parte do modelo SiST, apresentado nessa tese, foi publicado previamente em MORENO; TIMES; MATWIN (2015). As definições e os resultados adicionais foram submetidos no *International journal of geographical information science* (IJGIS), estando, portanto, aguardando o retorno da avaliação.

## 6.2 Trabalhos Futuros

Esta seção traz uma indicação de trabalhos em andamento ou a serem realizados com o que foi proposto nessa tese.

1. Executar algoritmos de ranqueamento sobre redes SiST com o objetivo de identificar usuários com poder de influência de encontros, usuários considerados líderes globais e usuários locais;
2. Correlacionar os resultados obtidos em 1 com os resultados da detecção de padrões de liderança de encontros;
3. Investigar a possibilidade de identificar eventos por meio das redes SiST e o padrão de grupo de encontros;
4. Investigar a possibilidade de identificar eventos por meio da execução de algoritmos de detecção de comunidades em redes SiST;
5. Analisar como é formada a componente principal de cada rede SiST utilizada nos experimentos desta tese;
6. Analisar as redes utilizadas nos experimentos de acordo com a evolução temporal das mesmas;
7. Investigar abordagens de desbalanceamento das redes para que as mesmas possam ser utilizadas em conjunto com modelos classificadores para tarefas de previsão, como previsão de laços sociais e previsão de encontros;

- ABRAMSON, D. et al. Node Assortativity in Complex Networks: an alternative approach. **Procedia Computer Science**, [S.l.], v.29, p.2449 – 2461, 2014.
- ADAMIC, L. A.; BUYUKKOKTEN, O.; ADAR, E. A social network caught in the Web. **First Monday**, [S.l.], v.8, n.6, 2003.
- AGARWAL, V.; BHARADWAJ, K. K. A collaborative filtering framework for friends recommendation in social networks based on interaction intensity and adaptive user similarity. **Social Netw. Analys. Mining**, [S.l.], v.3, n.3, p.359–379, 2013.
- ALBERT, R.; JEONG, H.; BARABASI, A. Error and attack tolerance of complex networks. **Nature**, [S.l.], v.406, n.6794, p.378–382, 2000.
- ALLAMANIS, M.; SCELLATO, S.; MASCOLO, C. Evolution of a Location-based Online Social Network: analysis and models. In: ACM CONFERENCE ON INTERNET MEASUREMENT CONFERENCE, 2012., NY, USA. **Proceedings...** ACM, 2012. p.145–158.
- AMAZON, A. **Alexa Internet, Inc.** Acessado em 17 de julho de 2016, <http://www.alexa.com/topsites/>.
- BACKSTROM, L.; LESKOVEC, J. Supervised random walks: predicting and recommending links in social networks. In: ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING, New York, NY, USA. **Proceedings...** ACM, 2011. p.635–644. (WSDM '11).
- BACKSTROM, L.; SUN, E.; MARLOW, C. Find Me if You Can: improving geographical prediction with social and spatial proximity. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 19., New York, NY, USA. **Proceedings...** ACM, 2010. p.61–70.
- BAKSHY, E. et al. Everyone's an Influencer: quantifying influence on twitter. In: FOURTH ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING, New York, NY, USA. **Proceedings...** ACM, 2011. p.65–74. (WSDM '11).
- BAO, J. et al. A survey on recommendations in location-based social networks. **GeoInformatica**, [S.l.], 2013.
- BARABASI, A.-L.; ALBERT, R. Emergence of Scaling in Random Networks. **Science**, [S.l.], v.286, n.5439, p.509–512, 1999.
- BENEVENUTO, F. et al. Characterizing User Behavior in Online Social Networks. In: ACM SIGCOMM CONFERENCE ON INTERNET MEASUREMENT CONFERENCE, 9., New York, NY, USA. **Proceedings...** ACM, 2009. p.49–62. (IMC '09).
- BERMINGHAM, A.; SMEATON, A. F. Classifying Sentiment in Microblogs: is brevity an advantage? In: ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 19., New York, NY, USA. **Proceedings...** ACM, 2010. p.1833–1836.
- BOCCALETTI, S. et al. Complex Networks : structure and dynamics. **Phys. Rep.**, [S.l.], v.424, n.4-5, p.175–308, Fervier 2006.

BONACICH, P.; LLOYD, P. Eigenvector-like measures of centrality for asymmetric relations. **Social Networks**, [S.l.], v.23, n.3, p.191–201, July 2001.

BORA, N. et al. Gang Networks, Neighborhoods and Holidays: spatiotemporal patterns in social media. In: SOCIALCOM. **Anais...** IEEE Computer Society, 2013. p.93–101.

BORGATTI, S. P. Centrality and network flow. **Social Networks**, [S.l.], v.27, n.1, p.55–71, 2005.

BOYD, D.; GOLDBERGER, S.; LOTAN, G. Tweet, Tweet, Retweet: conversational aspects of retweeting on twitter. In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 2010., Washington, USA. **Proceedings...** IEEE Computer Society, 2010. p.1–10.

BOYD, D. M.; ELLISON, N. B. Social network sites: definition, history, and scholarship. **Journal of Computer-Mediated Communication**, [S.l.], v.13, n.1, 2007.

BRILHANTE, I. R. et al. Trajectory Data Analysis Using Complex Networks. In: SYMPOSIUM ON INTERNATIONAL DATABASE ENGINEERING & APPLICATIONS, 15., New York, NY, USA. **Proceedings...** ACM, 2011. p.17–25. (IDEAS '11).

BROWN, C. et al. A place-focused model for social networks in cities. **CoRR**, [S.l.], v.abs/1308.2565, 2013.

BRZEZIŃSKI, M. Do wealth distributions follow power laws? Evidence from 'rich lists'. **Physica A: Statistical Mechanics and its Applications**, [S.l.], v.406, n.C, p.155–162, 2014.

BUCKLEY, F.; HARARY, F. **Distance in Graphs**. [S.l.]: Addison-Wesley, 1990.

CAVERLEE, J. et al. Public checkins versus private queries: measuring and evaluating spatial preference. In: INTERNATIONAL WORKSHOP ON LOCATION-BASED SOCIAL NETWORKS, 5., New York, NY, USA. **Proceedings...** ACM, 2012. p.40–47. (LBSN '12).

CHA, M. et al. Measuring user influence in twitter: the million follower fallacy. In: INTERNATIONAL AAAI CONFERENCE ON WEBLOGS AND SOCIAL MEDIA (ICWSM), 4. **Anais...** [S.l.: s.n.], 2010.

CHA, M.; MISLOVE, A.; GUMMADI, K. P. A Measurement-driven Analysis of Information Propagation in the Flickr Social Network. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 18., New York, NY, USA. **Proceedings...** ACM, 2009. p.721–730. (WWW '09).

CHAKRABARTI, S. et al. Automatic Resource Compilation by Analyzing Hyperlink Structure and Associated Text. **Computer Networks**, [S.l.], v.30, n.1-7, p.65–74, 1998.

CHENG, Z. et al. Exploring Millions of Footprints in Location Sharing Services. In: ICWSM. **Anais...** The AAAI Press, 2011.

CHO, E.; MYERS, S. A.; LESKOVEC, J. Friendship and mobility: user movement in location-based social networks. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 17., New York, NY, USA. **Proceedings...** ACM, 2011. p.1082–1090. (KDD '11).

CLARKE, K. A. A Simple Distribution-Free Test for Nonnested Model Selection. **Political Analysis**, [S.l.], v.15, n.3, p.347–363, 2007.

- CLAUSET, A.; SHALIZI, C. R.; NEWMAN, M. E. J. Power-Law Distributions in Empirical Data. **SIAM Rev.**, Philadelphia, PA, USA, v.51, n.4, p.661–703, Nov. 2009.
- CNN. **Facebook introduces check-in feature**. Acessado em 28 de junho de 2016, <http://edition.cnn.com/2010/TECH/social.media/08/18/facebook.location/>.
- COMPEAU, P.; PEVZNER, P.; TESLER, G. How to apply de Bruijn graphs to genome assembly. **Nature biotechnology**, [S.l.], v.29, n.11, p.987–991, 2011.
- CORMEN, T. H. et al. **Introduction to Algorithms**. 2nd.ed. [S.l.]: McGraw-Hill Higher Education, 2001.
- CRANSHAW, J. et al. Bridging the Gap Between Physical Location and Online Social Networks. In: ACM INTERNATIONAL CONFERENCE ON UBIQUITOUS COMPUTING, 12., New York, NY, USA. **Proceedings...** ACM, 2010. p.119–128. (UbiComp '10).
- DIESTEL, R. **Graph Theory**. [S.l.]: Springer, 1997. n.173. (Graduate Texts in Mathematics).
- DONG, C.; JIN, H.; KNIJENBURG, B. P. Predicting Privacy Behavior on Online Social Networks. In: NINTH INTERNATIONAL CONFERENCE ON WEB AND SOCIAL MEDIA, ICWSM 2015, UNIVERSITY OF OXFORD, OXFORD, UK, MAY 26-29, 2015. **Proceedings...** [S.l.: s.n.], 2015. p.91–100.
- EISENSTEIN, J. et al. A Latent Variable Model for Geographic Lexical Variation. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2010., Stroudsburg, PA, USA. **Proceedings...** Association for Computational Linguistics, 2010. p.1277–1287. (EMNLP '10).
- SHANKLIN, C. (Ed.). **Fundamentals of Database Systems**. 3rd.ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999.
- EMIRBAYER, M.; GOODWIN, J. Network Analysis, Culture, and the Problem of Agency. **American Journal of Sociology**, [S.l.], v.99, n.6, p.1411–1454, 1994.
- ERDÖS, P.; RÉNYI, A. On the Evolution of Random Graphs. In: PUBLICATION OF THE MATHEMATICAL INSTITUTE OF HUNGARIAN ACADEMY OF SCIENCES. **Anais...** [S.l.: s.n.], 1960. p.17–61.
- EULER, L. Solutio problematis ad geometriam situs pertinentis. **Commentarii Academiae Scientiarum Imperialis Petropolitanae**, [S.l.], v.8, p.128–140, 1736.
- FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. On Power-law Relationships of the Internet Topology. **SIGCOMM Comput. Commun. Rev.**, New York, NY, USA, v.29, n.4, p.251–262, Aug. 1999.
- FELDMAN, R. Techniques and Applications for Sentiment Analysis. **Commun. ACM**, New York, NY, USA, v.56, n.4, p.82–89, Apr. 2013.
- FERRARI, L. et al. Extracting urban patterns from location-based social networks. In: ACM SIGSPATIAL INTERNATIONAL WORKSHOP ON LOCATION-BASED SOCIAL NETWORKS, 3., New York, NY, USA. **Proceedings...** ACM, 2011. p.9–16. (LBSN '11).

- FERREIRA FILHO, J. a. B.; ACHER, M.; BARAIS, O. Challenges on Software Unbundling: growing and letting go. In: COMPANION PROCEEDINGS OF THE 14TH INTERNATIONAL CONFERENCE ON MODULARITY, New York, NY, USA. **Anais...** ACM, 2015. p.43–46.
- FREEMAN, L. C. Centrality in social networks conceptual clarification. **Social Networks**, [S.l.], p.215, 1978.
- GEORGIEV, P.; NOULAS, A.; MASCOLO, C. The Call of the Crowd: event participation in location-based social services. **CoRR**, [S.l.], v.abs/1403.7657, 2014.
- GEORGIEV, P.; NOULAS, A.; MASCOLO, C. Where Businesses Thrive: predicting the impact of the olympic games on local retailers through location-based services data. **CoRR**, [S.l.], v.abs/1403.7654, 2014.
- GILLESPIE, C. S. Fitting Heavy Tailed Distributions: the powerLaw package. **Journal of Statistical Software**, [S.l.], v.64, n.2, p.1–16, 2015.
- GONZALEZ, M. C.; HIDALGO, C. A.; BARABASI, A.-L. Understanding individual human mobility patterns. **Nature**, [S.l.], v.453, n.7196, p.779–782, June 2008.
- GROSS, R.; ACQUISTI, A. Information Revelation and Privacy in Online Social Networks. In: ACM WORKSHOP ON PRIVACY IN THE ELECTRONIC SOCIETY, 2005., New York, NY, USA. **Proceedings...** ACM, 2005. p.71–80. (WPES '05).
- GUZZO, R. A.; SHEA, G. P. Group performance and intergroup relations in organizations. In: DUNNETTE, M. D.; HOUGH, L. M. (Ed.). **Handbook of Industrial and Organizational Psychology**. 2nd.ed. Palo Alto, CA: Consulting Psychologists Press, 1992. v.3, p.269–313.
- HARARY, F. **Graph Theory**. [S.l.]: Addison Wesley, 1969.
- HOPPE, B.; REINELT, C. Social network analysis and the evaluation of leadership networks. **The Leadership Quarterly**, [S.l.], v.21, n.4, p.600–619, 2010.
- HORNIK, K. **R FAQ**. 2016.
- HUMPHREYS, L. Mobile Social Networks and Social Practice: a case study of dodgeball. **J. Computer-Mediated Communication**, [S.l.], v.13, n.1, p.341–360, 2007.
- JACKSON, M. O.; ROGERS, B. W. Meeting Strangers and Friends of Friends: how random are social networks? **American Economic Review**, [S.l.], v.97, n.3, p.890–915, June 2007.
- JIANG, J.; LI, C. Analyzing Social Event Participants for a Single Organizer. In: TENTH INTERNATIONAL CONFERENCE ON WEB AND SOCIAL MEDIA, COLOGNE, GERMANY, MAY 17-20, 2016. **Proceedings...** [S.l.: s.n.], 2016. p.599–602.
- KLEINBERG, J.; LAWRENCE, S. The Structure of the Web. **Science**, [S.l.], n.294, p.1849–1850, 2001.
- KLEINBERG, J. M. Authoritative Sources in a Hyperlinked Environment. **J. ACM**, New York, NY, USA, v.46, n.5, p.604–632, Sept. 1999.
- KRACKHARDT, D.; PORTER, L. W. The snowball effect: turnover embedded in communication networks. **Journal of Applied Psychology**, [S.l.], v.71, n.1, p.50–55, Feb. 1986.

- KUMAR, R. et al. Trawling the Web for Emerging Cyber-communities. In: EIGHTH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, New York, NY, USA. **Proceedings...** Elsevier North-Holland: Inc., 1999. p.1481–1493. (WWW '99).
- KURKA, D. B.; GODOY, A.; ZUBEN, F. J. V. Online Social Network Analysis: a survey of research applications in computer science. **CoRR**, [S.l.], v.abs/1504.05655, 2015.
- KYLASA, S. B.; KOLLIAS, G.; GRAMA, A. Social Ties and Checkin Sites: connections and latent structures in location based social networks. In: IEEE/ACM INTERNATIONAL CONFERENCE ON ADVANCES IN SOCIAL NETWORKS ANALYSIS AND MINING 2015, 2015., New York, NY, USA. **Proceedings...** ACM, 2015. p.194–201. (ASONAM '15).
- LEE, K. et al. SLAW: self-similar least-action human walk. **IEEE/ACM Trans. Netw.**, Piscataway, NJ, USA, v.20, n.2, p.515–529, Apr. 2012.
- LERMAN, K.; GHOSH, R. Information contagion: an empirical study of the spread of news on digg and twitter social networks. In: INTERNATIONAL CONFERENCE ON WEBLOGS AND SOCIAL MEDIA (ICWSM), 4. **Proceedings...** [S.l.: s.n.], 2010.
- LI, N.; CHEN, G. Analysis of a Location-Based Social Network. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND ENGINEERING - VOLUME 04, 2009., Washington, DC, USA. **Proceedings...** IEEE Computer Society, 2009. p.263–270.
- LICHTENWALTER, R.; LUSSIER, J. T.; CHAWLA, N. V. New perspectives and methods in link prediction. In: KDD. **Anais...** [S.l.: s.n.], 2010. p.243–252.
- LINDQVIST, J. et al. I'm the Mayor of My House: examining why people use foursquare - a social-driven location sharing application. In: SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, NY, USA. **Proceedings...** ACM, 2011. p.2409–2418.
- LINKEDIN. **A brief history of LinkedIn**. Acessado em 10 de julho de 2016, <https://ourstory.linkedin.com/>.
- MCPHERSON, M.; SMITH-LOVIN, L.; COOK, J. M. Birds of a Feather: homophily in social networks. **Annual Review of Sociology**, [S.l.], v.27, n.1, p.415–444, 2001.
- MERTON, R. K. The Matthew effect in science: the reward and communication systems of science are considered. **Science**, [S.l.], v.159, n.3810, p.56–63, 1968.
- MILGRAM, S. The Small World Problem. **Psychology Today**, [S.l.], v.67, n.1, p.61–67, 1967.
- MISLOVE, A. **Online Social Networks: measurement, analysis, and applications to distributed information systems**. 2009. Tese (Doutorado em Ciência da Computação) — Rice University, Department of Computer Science.
- MISLOVE, A. et al. Measurement and analysis of online social networks. In: ACM SIGCOMM CONFERENCE ON INTERNET MEASUREMENT, 7., New York, NY, USA. **Proceedings...** ACM, 2007. p.29–42. (IMC '07).
- MITCHELL, M. **Complexity: a guided tour**. New York, NY, USA: Oxford University Press, Inc., 2009.

MITZENMACHER, M. A Brief History of Generative Models for Power Law and Lognormal Distributions. **Internet Mathematics**, [S.l.], v.1, n.2, p.226–251, 2004.

MORENO, B. N.; TIMES, V. C.; MATWIN, S. A Spatio-temporal Network Model to Represent and Analyze LBSNs. In: INT. WORKSHOP ON PERVASIVE COLLABORATION AND SOCIAL NETWORKING, 6. **Anais...** IEEE, 2015. (PerCol).

MOYANO, L. G.; THOMAE, O. R. M.; FRÍAS-MARTÍNEZ, E. Uncovering the Spatio-temporal Structure of Social Networks Using Cell Phone Records. In: ICDM WORKSHOPS. **Anais...** IEEE Computer Society, 2012. p.242–249.

MUSICA, U. **SXSW 2010**: festival começa a se despedir do público com feira de guitarras. Acessado em 09 de junho de 2016, <http://goo.gl/fEgsZS>.

NEWMAN, M. E. J. Mixing patterns in networks. **Phys. Rev. E**, [S.l.], v.67, n.2, p.026126, Feb. 2003.

NEWMAN, M. E. J. The Structure and Function of Complex Networks. **SIAM Review**, [S.l.], v.45, n.2, p.167–256, 2003.

NOULAS, A. et al. An Empirical Study of Geographic User Activity Patterns in Foursquare. In: ICWSM. **Anais...** The AAAI Press, 2011.

NOULAS, A. et al. A tale of many cities: universal patterns in human urban mobility. **CoRR**, [S.l.], v.abs/1108.5355, 2011.

O'REILLY, T. **What Is Web 2.0 - Design Patterns and Business Models for the Next Generation of Software**. [S.l.]: Tim O'Reilly, 2005.

PAGE, L. et al. The PageRank citation ranking: bringing order to the web. In: INTERNATIONAL WORLD WIDE WEB CONFERENCE, 7., Brisbane, Australia. **Proceedings...** [S.l.: s.n.], 1998. p.161–172.

PANG, B.; LEE, L. Opinion Mining and Sentiment Analysis. **Found. Trends Inf. Retr.**, Hanover, MA, USA, v.2, n.1-2, p.1–135, Jan. 2008.

PARK, K.; WILLINGER, W. **Self-Similar Network Traffic and Performance Evaluation**. 1st.ed. New York, NY, USA: John Wiley & Sons, Inc., 2000.

PELECHRINIS, K.; KRISHNAMURTHY, P. Location Affiliation Networks: bonding social and spatial information. In: ECML/PKDD (2). **Anais...** Springer, 2012. p.531–547. (Lecture Notes in Computer Science, v.7524).

RICE, S. A. The Identification of Blocs in Small Political Bodies. **American Political Science Review**, [S.l.], v.21, p.619–627, 8 1927.

RIEDERER, C. J. et al. "I Don't Have a Photograph, but You Can Have My Footprints.": revealing the demographics of location data. In: ACM ON CONFERENCE ON ONLINE SOCIAL NETWORKS, 2015., New York, NY, USA. **Proceedings...** ACM, 2015. p.185–195.

ROGERS, R. Debanalizing Twitter: the transformation of an object of study. In: ANNUAL ACM WEB SCIENCE CONFERENCE, 5., NY, USA. **Proceedings...** ACM, 2013. p.356–365.

- SCELLATO, S. et al. Socio-Spatial Properties of Online Location-Based Social Networks. In: ICWSM. **Anais...** The AAAI Press, 2011.
- SCELLATO, S.; MASCOLO, C. Measuring User Activity on an Online Location-based Social Network. In: NETSCICOM'11. **Proceedings...** [S.l.: s.n.], 2011.
- SCELLATO, S.; NOULAS, A.; MASCOLO, C. Exploiting Place Features in Link Prediction on Location-based Social Networks. In: ACM SIGKDD, 17., NY, USA. **Proceedings...** ACM, 2011. p.1046–1054.
- SCHNEIDER, F. et al. Understanding Online Social Network Usage from a Network Perspective. In: ACM SIGCOMM CONFERENCE ON INTERNET MEASUREMENT CONFERENCE, 9., New York, NY, USA. **Proceedings...** ACM, 2009. p.35–48. (IMC '09).
- SESHADRI, M. et al. Mobile call graphs: beyond power-law and lognormal distributions. In: KDD. **Anais...** ACM, 2008. p.596–604.
- SKIENA, S. **Implementing discrete mathematics - combinatorics and graph theory with Mathematica.** [S.l.]: Addison-Wesley, 1990. I-VIIIp.
- STATISTA. **The Statistics Portal.** Acessado em 10 de julho de 2016, <http://www.statista.com/statistics/264810/>.
- STATISTA. **The Statistics Portal.** Acessado em 10 de julho de 2016, <http://www.statista.com/statistics/274050/>.
- TAN, C. et al. User-level Sentiment Analysis Incorporating Social Networks. In: ACM SIGKDD, 17., New York, NY, USA. **Proceedings...** ACM, 2011. p.1397–1405. (KDD '11).
- TWITTER. **Twitter Blog.** Acessado em 10 de julho de 2016, <https://blog.twitter.com/2012/twitter-turns-six>.
- VASCONCELOS, M. A. et al. Tips, Dones and Todos: uncovering user profiles in foursquare. In: FIFTH ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING, New York, NY, USA. **Proceedings...** ACM, 2012. p.653–662. (WSDM '12).
- VASCONCELOS, M. A. et al. Popularity dynamics of foursquare micro-reviews. In: COSN. **Anais...** ACM, 2014. p.119–130.
- VUONG, Q. H. Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses. **Econometrica**, [S.l.], v.57, n.2, p.307–33, 1989.
- WANG, D. et al. Human Mobility, Social Ties, and Link Prediction. In: ACM SIGKDD, 17., New York, NY, USA. **Proceedings...** ACM, 2011. p.1100–1108. (KDD '11).
- WATTS, D. **Six Degrees: the science of a connected age.** [S.l.]: W. W. Norton, 2004.
- WATTS, D. J. The "new" science of networks. **Annual Review of Sociology**, [S.l.], v.30, p.243–270, 2004.
- WATTS, D.; STROGATZ, S. Collective dynamics of 'small-world' networks. **Nature**, [S.l.], n.393, p.440–442, 1998.
- HALL, P. (Ed.). **Introduction to Graph Theory.** [S.l.]: Prentice Hall, 2001.

WILKEN, R. The de-gamification of Foursquare? **Social, Casual and Mobile Games: The Changing Gaming Landscape**, [S.l.], p.5, 2016.

WILSON, C. et al. User Interactions in Social Networks and Their Implications. In: ACM EUROPEAN CONFERENCE ON COMPUTER SYSTEMS, 4., New York, NY, USA. **Proceedings...** ACM, 2009. p.205–218. (EuroSys '09).

WU, S. et al. Who Says What to Whom on Twitter. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 20., NY, USA. **Proceedings...** ACM, 2011. p.705–714.

YIP, M.; SHADBOLT, N.; WEBBER, C. Structural analysis of online criminal social networks. In: ISI. **Anais...** [S.l.: s.n.], 2012.

YOOK, S.-H.; RADICCHI, F.; MEYER-ORTMANN, H. Self-similar scale-free networks and disassortativity. **Phys. Rev. E**, [S.l.], v.72, p.045105, Oct 2005.

YUAN, Q.; CONG, G.; SUN, A. Graph-based Point-of-interest Recommendation with Geographical and Temporal Influences. In: CIKM, 23., NY, USA. **Proceedings...** ACM, 2014. p.659–668.

YUAN, Q. et al. Time-aware Point-of-interest Recommendation. In: INTERNATIONAL ACM SIGIR, 36. **Proceedings...** [S.l.: s.n.], 2013. p.363–372.

ZHUANG, H. et al. Inferring Geographic Coincidence in Ephemeral Social Networks. In: EUROPEAN CONFERENCE ON MACHINE LEARNING AND KNOWLEDGE DISCOVERY IN DATABASES, 2012., Berlin, Heidelberg. **Proceedings...** Springer-Verlag, 2012. p.613–628.

# **Apêndice**

## A

## Tabelas Adicionais do Teste de Vuong do Capítulo 4

Base	$\Delta t$	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-\ln}$	$p_{pl-\ln}$	Resultado
BKT_CHO	15 min	<b>0.98</b>	0.29	0.76	Inconclusivo
BKT_CHO	30 min	0.00	-1.55	0.12	Log-normal
BKT_CHO	1h	0.00	-4.53	<b>0.00</b>	Log-normal
BKT_CHO	2h	<b>0.84</b>	0.54	<b>0.00</b>	Inconclusivo
GOW_CHO	15 min	0.00	8.66	<b>0.00</b>	Log-normal
GOW_CHO	30 min	0.00	-12.18	<b>0.00</b>	Log-normal
GOW_CHO	1h	0.00	17.49	<b>0.00</b>	Log-normal
GOW_CHO	2h	0.00	-1.25	0.20	Log-normal
GOW_SCE	15 min	0.00	-12.31	<b>0.00</b>	Log-normal
GOW_SCE	30 min	0.00	18.23	<b>0.00</b>	Log-normal
GOW_SCE	1h	0.00	37.72	<b>0.00</b>	Log-normal
GOW_SCE	2h	0.00	43.04	<b>0.00</b>	Log-normal
4QR_YUA	15 min	<b>0.44</b>	-4.46	<b>0.00</b>	Power-law
4QR_YUA	30 min	<b>0.49</b>	-4.69	<b>0.00</b>	Power-law
4QR_YUA	1h	<b>0.14</b>	-4.36	<b>0.00</b>	Power-law
4QR_YUA	2h	<b>0.28</b>	-1.53	0.12	Power-law
TWT_EIN	15 min	<b>0.37</b>	-1.23	0.21	Log-normal
TWT_EIN	30 min	0.02	-2.11	<b>0.03</b>	Log-normal
TWT_EIN	1h	<b>0.28</b>	-0.49	0.61	Log-normal
TWT_EIN	2h	<b>0.35</b>	-0.69	0.48	Log-normal
TWT_CHE	15 min	0.00	-13.00	<b>0.00</b>	Power-law
TWT_CHE	30 min	0.00	-13.87	<b>0.00</b>	Power-law
TWT_CHE	1h	0.00	-10.97	<b>0.00</b>	Log-normal
TWT_CHE	2h	0.00	-13.48	<b>0.00</b>	Log-normal

**Tabela A.1:** Teste de Vuong para o número de encontros não sociais distribuídos por pares de usuários.

Base	$\Delta t$	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	Resultado
BKT_CHO	15 min	0.00	3.03	<b>0.00</b>	Inconclusivo
BKT_CHO	30 min	0.00	-3.45	<b>0.00</b>	Log-normal
BKT_CHO	1h	0.00	-3.50	<b>0.00</b>	Log-normal
BKT_CHO	2h	0.00	3.69	<b>0.00</b>	Inconclusivo
GOW_CHO	15 min	0.00	-9.30	<b>0.00</b>	Log-normal
GOW_CHO	30 min	0.00	-8.03	<b>0.00</b>	Log-normal
GOW_CHO	1h	0.00	-7.12	<b>0.00</b>	Log-normal
GOW_CHO	2h	0.00	-6.52	<b>0.00</b>	Log-normal
GOW_SCE	15 min	0.00	-10.36	<b>0.00</b>	Log-normal
GOW_SCE	30 min	0.00	-6.39	<b>0.00</b>	Log-normal
GOW_SCE	1h	0.00	-4.03	<b>0.00</b>	Log-normal
GOW_SCE	2h	0.00	-3.31	<b>0.00</b>	Log-normal
4QR_YUA	15 min	<b>0.26</b>	-0.82	0.40	Power-law
4QR_YUA	30 min	<b>0.43</b>	-1.43	0.15	Power-law
4QR_YUA	1h	<b>0.13</b>	-1.29	0.19	Power-law
4QR_YUA	2h	<b>0.38</b>	-1.12	0.26	Power-law
TWT_EIN	15 min	0.00	-5.09	<b>0.00</b>	Log-normal
TWT_EIN	30 min	0.01	-3.89	<b>0.00</b>	Log-normal
TWT_EIN	1h	0.00	-3.22	<b>0.00</b>	Log-normal
TWT_EIN	2h	0.00	-3.88	<b>0.00</b>	Log-normal
TWT_CHE	15 min	<b>0.69</b>	-0.91	0.36	Power-law
TWT_CHE	30 min	<b>0.57</b>	-0.93	0.35	Power-law
TWT_CHE	1h	0.00	-10.61	<b>0.00</b>	Log-normal
TWT_CHE	2h	<b>1.00</b>	-0.90	0.36	Power-law

**Tabela A.2:** Teste de Vuong para o número de encontros não sociais distribuídos por usuários precursores.

Base	$\Delta t$	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	Resultado
BKT_CHO	15 min	0.00	5.12	<b>0.00</b>	Inconclusivo
BKT_CHO	30 min	0.00	-4.14	<b>0.00</b>	Log-normal
BKT_CHO	1h	0.00	-3.08	<b>0.00</b>	Log-normal
BKT_CHO	2h	0.00	4.29	<b>0.00</b>	Log-normal
GOW_CHO	15 min	0.00	-9.06	<b>0.00</b>	Inconclusivo
GOW_CHO	30 min	0.00	-7.89	<b>0.00</b>	Log-normal
GOW_CHO	1h	0.00	-7.13	<b>0.00</b>	Log-normal
GOW_CHO	2h	0.00	-6.16	<b>0.00</b>	Log-normal
GOW_SCE	15 min	0.00	-10.41	<b>0.00</b>	Log-normal
GOW_SCE	30 min	0.00	-6.53	<b>0.00</b>	Log-normal
GOW_SCE	1h	0.00	-4.21	<b>0.00</b>	Log-normal
GOW_SCE	2h	0.00	-3.28	<b>0.00</b>	Log-normal
4QR_YUA	15 min	<b>0.44</b>	-1.36	0.17	Power-law
4QR_YUA	30 min	<b>0.30</b>	-1.26	0.20	Power-law
4QR_YUA	1h	0.02	-2.39	<b>0.01</b>	Log-normal
4QR_YUA	2h	<b>0.15</b>	-0.86	0.38	Power-law
TWT_EIN	15 min	0.00	-4.98	<b>0.00</b>	Log-normal
TWT_EIN	30 min	0.00	-3.46	<b>0.00</b>	Log-normal
TWT_EIN	1h	0.00	-2.76	<b>0.00</b>	Log-normal
TWT_EIN	2h	0.00	-2.72	<b>0.00</b>	Log-normal
TWT_CHE	15 min	0.00	-16.08	<b>0.00</b>	Log-normal
TWT_CHE	30 min	<b>0.45</b>	-0.98	0.32	Power-law
TWT_CHE	1h	<b>0.47</b>	-0.82	0.40	Power-law
TWT_CHE	2h	<b>1.00</b>	-0.97	0.32	Power-law

**Tabela A.3:** Teste de Vuong para o número de encontros não sociais distribuídos por usuários seguidores.

Base	$\Delta t$	$p_{pl}$	$\mathcal{L} \mathcal{R}_{pl-ln}$	$p_{pl-ln}$	Resultado
BKT_CHO	15 min	0.01	1.34	0.17	Inconclusivo
BKT_CHO	30 min	0.00	1.18	0.23	Inconclusivo
BKT_CHO	1h	<b>0.38</b>	1.08	0.27	Power-law
BKT_CHO	2h	0.01	-1.43	0.15	Inconclusivo
GOW_CHO	15 min	0.00	-4.63	<b>0.00</b>	Log-normal
GOW_CHO	30 min	<b>0.75</b>	-0.53	0.59	Power-law
GOW_CHO	1h	<b>0.93</b>	-0.55	0.58	Power-law
GOW_CHO	2h	0.06	0.73	0.46	Inconclusivo
GOW_SCE	15 min	0.00	13.42	<b>0.00</b>	Inconclusivo
GOW_SCE	30 min	0.00	13.75	<b>0.00</b>	Inconclusivo
GOW_SCE	1h	0.00	14.32	<b>0.00</b>	Inconclusivo
GOW_SCE	2h	<b>0.71</b>	-0.38	0.69	Inconclusivo
4QR_YUA	15 min	0.00	-2.84	<b>0.00</b>	Log-normal
4QR_YUA	30 min	0.00	-2.76	<b>0.00</b>	Log-normal
4QR_YUA	1h	0.02	-2.24	<b>0.02</b>	Log-normal
4QR_YUA	2h	0.00	-3.29	<b>0.00</b>	Log-normal
TWT_EIN	15 min	0.08	1.56	0.11	Log-normal
TWT_EIN	30 min	<b>0.38</b>	1.25	0.20	Inconclusivo
TWT_EIN	1h	<b>0.87</b>	0.81	0.41	Power-law
TWT_EIN	2h	<b>0.31</b>	0.81	0.41	Power-law
TWT_CHE	15 min	<b>0.96</b>	-1.32	0.18	Power-law
TWT_CHE	30 min	<b>0.70</b>	-1.80	<b>0.07</b>	Inconclusivo
TWT_CHE	1h	0.00	-6.49	<b>0.00</b>	Log-normal
TWT_CHE	2h	0.00	-7.13	<b>0.00</b>	Log-normal

**Tabela A.4:** Teste de Vuong para o número de encontros não sociais distribuídos por locais.

## B

## Propriedades das Redes SiST

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	326	763	4.680982	11	0.5761586	3.430506	0.1438758
<b>Seg</b>	277	723	5.220217	18	0.5363307	4.384867	0.1878786
<b>Ter</b>	256	545	4.257812	31	0.5773688	4.494185	0.1982578
<b>Qua</b>	222	236	2.126126	17	0.5976828	3.980611	0.4395269
<b>Qui</b>	168	129	1.535714	11	0.5574501	1.376344	0.3016094
<b>Sex</b>	305	569	3.731148	18	0.5371661	4.852488	0.1977351
<b>Sab</b>	408	798	3.911765	16	0.53352	4.764726	0.1550702

Tabela B.1: Redes SiST não sociais de BKT\_CHO para  $\Delta t = 15$  min.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	395	1292	6.541772	19	0.6258478	4.07363	0.22855
<b>Seg</b>	347	1239	7.14121	14	0.6188319	3.536028	0.183235
<b>Ter</b>	330	917	5.557576	33	0.645313	3.954623	0.2425843
<b>Qua</b>	290	403	2.77931	24	0.6255132	4.825564	0.3552406
<b>Qui</b>	228	210	1.842105	16	0.6383333	1.494286	0.203151
<b>Sex</b>	378	993	5.253968	15	0.5838057	4.046009	0.2424917
<b>Sab</b>	477	1387	5.815514	22	0.598857	3.854454	0.1621658

Tabela B.2: Redes SiST não sociais de BKT\_CHO para  $\Delta t = 30$  min.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	467	2079	8.90364	20	0.6635414	3.496439	0.2650765
<b>Seg</b>	416	1942	9.336538	13	0.6531854	3.120782	0.2273196
<b>Ter</b>	406	1476	7.270936	51	0.6394555	3.669632	0.2870639
<b>Qua</b>	374	716	3.828877	41	0.6593722	4.40841	0.3219469
<b>Qui</b>	299	332	2.220736	9	0.6578937	1.765487	0.2177184
<b>Sex</b>	468	1715	7.32906	14	0.6472385	3.390273	0.2512387
<b>Sab</b>	568	2373	8.355634	16	0.679495	3.463496	0.1895998

Tabela B.3: Redes SiST não sociais de BKT\_CHO para  $\Delta t = 1$  hora.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	535	3202	11.97009	16	0.7068143	3.085646	0.3031089
<b>Seg</b>	498	2907	11.6747	12	0.6879356	2.80211	0.2830761
<b>Ter</b>	473	2226	9.412262	97	0.7007143	3.179516	0.3758284
<b>Qua</b>	441	1134	5.142857	21	0.7022073	3.938261	0.3968396
<b>Qui</b>	352	518	2.943182	11	0.7174297	1.858917	0.4244155
<b>Sex</b>	569	2818	9.905097	20	0.6754584	3.154182	0.3111714
<b>Sab</b>	349	491	2.813754	14	0.6980062	2.132645	0.4720843

**Tabela B.4:** Redes SiST não sociais de BKT\_CHO para  $\Delta t = 2$  horas.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	138	138	2	17	0.3864286	3.290462	0.1493919
<b>Seg</b>	95	113	2.378947	14	0.3862434	3.445104	0.1686381
<b>Ter</b>	98	116	2.367347	35	0.4281994	2.696774	0.2011859
<b>Qua</b>	73	59	1.616438	7	0.8050265	1.494845	0.2891925
<b>Qui</b>	65	44	1.353846	8	0.75	1.06383	0.5654528
<b>Sex</b>	109	103	1.889908	14	0.155452	3.926531	0.1244286
<b>Sab</b>	171	146	1.707602	10	0.3039479	2.743982	0.1505745

**Tabela B.5:** Redes SiST sociais de BKT\_CHO para  $\Delta t = 15$  min.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	158	192	2.43038	17	0.4215862	3.73384	0.02494558
<b>Seg</b>	114	160	2.807018	17	0.4713879	2.99799	0.2675491
<b>Ter</b>	129	161	2.496124	53	0.3817402	3.093591	0.1699232
<b>Qua</b>	103	82	1.592233	16	0.6052381	1.779874	0.3728226
<b>Qui</b>	75	53	1.413333	8	0.75	1.053571	0.4365006
<b>Sex</b>	150	161	2.146667	16	0.337185	3.893886	0.1820859
<b>Sab</b>	195	203	2.082051	17	0.4926686	3.290404	0.1213149

**Tabela B.6:** Redes SiST sociais de BKT\_CHO para  $\Delta t = 30$  min.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	184	253	2.75	20	0.4875056	3.265388	0.08979933
<b>Seg</b>	142	205	2.887324	25	0.5194625	3.369935	0.2026309
<b>Ter</b>	154	209	2.714286	114	0.472487	2.952599	0.155102
<b>Qua</b>	130	116	1.784615	19	0.6466184	2.29375	0.2809352
<b>Qui</b>	90	64	1.422222	8	0.75	1.058824	0.1483204
<b>Sex</b>	178	222	2.494382	26	0.3940942	3.595531	0.1798837
<b>Sab</b>	214	257	2.401869	22	0.5255869	3.184524	0.1302301

**Tabela B.7:** Redes SiST sociais de BKT\_CHO para  $\Delta t = 1$  hora.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	215	326	3.032558	21	0.4773771	3.263128	0.1456928
<b>Seg</b>	170	248	2.917647	34	0.5033604	3.320531	0.2148947
<b>Ter</b>	181	260	2.872928	213	0.5047797	2.908028	0.1394859
<b>Qua</b>	153	140	1.830065	23	0.6284539	2.391827	0.2732668
<b>Qui</b>	107	79	1.476636	8	0.4615385	1.131868	0.06511655
<b>Sex</b>	224	318	2.839286	38	0.4895785	3.088796	0.2729223
<b>Sab</b>	119	109	1.831933	13	0.7314286	1.170543	0.6305268

**Tabela B.8:** Redes SiST sociais de BKT\_CHO para  $\Delta t = 2$  horas.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	1802	39694	44.05549	21	0.5188858	2.823418	0.0995551
<b>Seg</b>	1580	41358	52.3519	30	0.5524467	2.810591	0.2041945
<b>Ter</b>	1510	28094	37.2106	18	0.549769	3.040879	0.1648129
<b>Qua</b>	1097	4767	8.690975	15	0.5625095	4.337427	0.4616402
<b>Qui</b>	698	842	2.412607	27	0.5140348	6.482826	0.4068739
<b>Sex</b>	1332	20662	31.02402	12	0.5593179	2.958679	0.1847567
<b>Sab</b>	1726	37326	43.25145	9	0.5243046	2.800261	0.1057499

**Tabela B.9:** Redes SiST não sociais de GOW\_CHO para  $\Delta t = 15$  min.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	1949	69247	71.059	24	0.5703234	2.576868	0.06681544
<b>Seg</b>	1695	71460	84.31858	32	0.6062523	2.564191	0.178521
<b>Ter</b>	1623	49348	60.81084	17	0.6027405	2.728959	0.1324758
<b>Qua</b>	1265	8796	13.90672	12	0.5934135	3.602901	0.4289114
<b>Qui</b>	870	1462	3.36092	27	0.5327499	4.992534	0.4399675
<b>Sex</b>	1459	38972	53.42289	14	0.5953529	2.726945	0.1761522
<b>Sab</b>	1864	66800	71.67382	8	0.5700409	2.548081	0.07648365

**Tabela B.10:** Redes SiST não sociais de GOW\_CHO para  $\Delta t = 30$  min.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	2089	110176	105.482	24	0.6200239	2.419795	0.01437838
<b>Seg</b>	1831	109862	120.0022	32	0.6516471	2.424336	0.124215
<b>Ter</b>	1751	79016	90.25243	17	0.6577846	2.500827	0.08831139
<b>Qua</b>	1419	15818	22.29457	14	0.6330209	3.319639	0.4064291
<b>Qui</b>	1027	2525	4.917235	27	0.5842374	3.985586	0.3702383
<b>Sex</b>	1578	71743	90.92902	17	0.6493241	2.72077	0.1367089
<b>Sab</b>	2034	113213	111.3206	9	0.6263617	2.358789	0.04417002

**Tabela B.11:** Redes SiST não sociais de GOW\_CHO para  $\Delta t = 1$  hora.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	996	10732	21.5502	12	0.7024172	3.06785	0.2331116
<b>Seg</b>	1982	151345	152.7195	32	0.689342	2.295458	0.03890404
<b>Ter</b>	1894	116742	123.2756	17	0.6920264	2.383524	0.05391883
<b>Qua</b>	1604	26867	33.5	15	0.6742265	3.165423	0.3540476
<b>Qui</b>	1204	4263	7.081395	27	0.6414984	3.730259	0.3359276
<b>Sex</b>	1723	126202	146.491	15	0.7169186	2.67554	0.09308806
<b>Sab</b>	2220	185047	166.709	9	0.6878762	2.230291	0.03307212

**Tabela B.12:** Redes SiST não sociais de GOW\_CHO para  $\Delta t = 2$  horas.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	941	1173	2.493092	30	0.3701853	7.926902	0.2402625
<b>Seg</b>	760	951	2.502632	30	0.3660597	6.243659	0.1572116
<b>Ter</b>	667	697	2.089955	22	0.357242	5.076494	0.3910226
<b>Qua</b>	504	411	1.630952	12	0.5956867	2.159675	0.5112316
<b>Qui</b>	331	206	1.244713	27	0.6266667	1.081081	0.4130257
<b>Sex</b>	693	631	1.821068	19	0.2999997	3.585435	0.2941683
<b>Sab</b>	888	1086	2.445946	28	0.3897269	6.125028	0.05080373

**Tabela B.13:** Redes SiST sociais de GOW\_CHO para  $\Delta t = 15$  min.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	1091	1603	2.938588	25	0.3543462	6.910822	0.1649699
<b>Seg</b>	880	1340	3.045455	32	0.3639917	5.820034	0.0972121
<b>Ter</b>	778	957	2.460154	22	0.3286042	6.355008	0.3682775
<b>Qua</b>	577	506	1.753899	21	0.5030748	3.4049	0.5228613
<b>Qui</b>	396	256	1.292929	27	0.6969697	1.096774	0.4420466
<b>Sex</b>	826	917	2.220339	46	0.3113443	7.199976	0.2244824
<b>Sab</b>	1011	1538	3.042532	36	0.3663394	5.246966	0.01482546

**Tabela B.14:** Redes SiST sociais de GOW\_CHO para  $\Delta t = 30$  min.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	1205	2137	3.546888	24	0.3449201	5.717574	0.06763101
<b>Seg</b>	977	1779	3.64176	32	0.3563844	5.346311	0.0777511
<b>Ter</b>	879	1328	3.021615	22	0.3490932	5.683323	0.2475446
<b>Qua</b>	661	667	2.018154	20	0.5098358	5.555778	0.4647329
<b>Qui</b>	442	302	1.366516	27	0.657619	1.19888	0.3649078
<b>Sex</b>	924	1334	2.887446	43	0.3117577	5.814233	0.1887841
<b>Sab</b>	1105	2108	3.815385	36	0.3626949	4.706791	-0.01707701

**Tabela B.15:** Redes SiST sociais de GOW\_CHO para  $\Delta t = 1$  hora.

<b>Rede</b>	<b>nós</b>	<b>conexões</b>	<b>grau</b>	<b>diâmetro</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>assortatividade</b>
<b>Dom</b>	492	510	2.073171	26	0.4829083	5.702393	0.4472144
<b>Seg</b>	1084	2301	4.245387	32	0.3560246	4.802201	0.01662308
<b>Ter</b>	986	1752	3.553753	25	0.340791	5.500892	0.1634766
<b>Qua</b>	734	828	2.256131	30	0.5001909	6.696123	0.3992826
<b>Qui</b>	508	366	1.440945	27	0.620915	1.691011	0.3417434
<b>Sex</b>	1014	1935	3.816568	41	0.3192111	5.108581	0.1088637
<b>Sab</b>	1216	2832	4.657895	37	0.361629	4.195079	-0.05015286

**Tabela B.16:** Redes SiST sociais de GOW\_CHO para  $\Delta t = 2$  horas.

<b>Rede</b>	<b>nós</b>	<b>conexões</b>	<b>grau</b>	<b>diâmetro</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>assortatividade</b>
<b>Dom</b>	4531	130733	57.70603	43	0.5738852	2.929207	0.1947359
<b>Seg</b>	3771	130360	69.13816	30	0.5778109	3.137809	0.3260197
<b>Ter</b>	3609	93628	51.88584	19	0.5908423	3.162341	0.2314842
<b>Qua</b>	3252	16235	9.984625	27	0.6052673	5.141916	0.4916302
<b>Qui</b>	2633	4703	3.572351	50	0.6431537	8.46083	0.4423549
<b>Sex</b>	3325	49301	29.65474	36	0.5751143	3.688869	0.2381195
<b>Sab</b>	4377	70325	32.13388	46	0.5550736	3.1938	0.2775638

**Tabela B.17:** Redes SiST não sociais de GOW\_SCE para  $\Delta t = 15$  min.

<b>Rede</b>	<b>nós</b>	<b>conexões</b>	<b>grau</b>	<b>diâmetro</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>assortatividade</b>
<b>Dom</b>	4933	202605	82.14271	44	0.6020963	2.80909	0.140443
<b>Seg</b>	4084	220707	108.0837	32	0.6291843	3.476558	0.2975007
<b>Ter</b>	3957	163136	82.45438	22	0.6283641	3.104653	0.1993362
<b>Qua</b>	3629	30544	16.83329	25	0.6392707	4.677686	0.4851914
<b>Qui</b>	3051	8667	5.681416	39	0.6707996	8.20006	0.4481633
<b>Sex</b>	3619	93243	51.5297	36	0.609749	3.372696	0.2263574
<b>Sab</b>	4822	124559	51.6628	46	0.5861683	3.842787	0.2733625

**Tabela B.18:** Redes SiST não sociais de GOW\_SCE para  $\Delta t = 30$  min.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	5320	304614	114.5165	46	0.6395462	3.149461	0.1175861
<b>Seg</b>	4410	327087	148.3388	35	0.6604566	3.588906	0.2589869
<b>Ter</b>	4285	263691	123.0763	22	0.6695502	3.234525	0.1655397
<b>Qua</b>	3999	55849	27.93148	17	0.6736857	4.209865	0.4789671
<b>Qui</b>	3495	16076	9.199428	36	0.6803236	7.368547	0.4681675
<b>Sex</b>	3942	173105	87.82598	36	0.6404476	3.354768	0.2192671
<b>Sab</b>	5218	206037	78.97164	51	0.6228509	3.715643	0.2617559

**Tabela B.19:** Redes SiST não sociais de GOW\_SCE para  $\Delta t = 30$  min.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	3122	68162	43.6656	29	0.7364637	4.81001	0.4364306
<b>Seg</b>	4756	432482	181.868	32	0.683351	3.282104	0.2230429
<b>Ter</b>	4608	384050	166.6884	25	0.701414	3.304552	0.1569968
<b>Qua</b>	4386	98658	44.98769	17	0.7009154	3.934012	0.4677876
<b>Qui</b>	3984	29297	14.70733	33	0.7059864	7.077231	0.4766978
<b>Sex</b>	4338	311839	143.7709	36	0.684952	3.518337	0.2316269
<b>Sab</b>	5712	319967	112.0333	73	0.647865	3.696131	0.2002569

**Tabela B.20:** Redes SiST não sociais de GOW\_SCE para  $\Delta t = 2$  horas.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	2397	3665	3.057989	48	0.4655326	6.527082	0.2138868
<b>Seg</b>	1761	2680	3.043725	47	0.4256815	6.379106	0.3230196
<b>Ter</b>	1634	2188	2.678091	39	0.4486344	6.776835	0.3534753
<b>Qua</b>	1283	1131	1.763055	22	0.5936731	4.27221	0.510379
<b>Qui</b>	973	675	1.387461	33	0.7316302	1.122667	0.5022937
<b>Sex</b>	1482	1526	2.059379	37	0.5283797	5.93021	0.4421564
<b>Sab</b>	2094	2454	2.34384	46	0.5386232	7.024827	0.3198488

**Tabela B.21:** Redes SiST sociais de GOW\_SCE para  $\Delta t = 15$  min.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	2618	4530	3.460657	53	0.44674	5.862989	0.2226922
<b>Seg</b>	1938	3579	3.693498	52	0.4170728	5.59604	0.2720181
<b>Ter</b>	1818	2890	3.179318	44	0.4209861	6.526655	0.3066641
<b>Qua</b>	1419	1382	1.947851	28	0.5467802	6.152371	0.5230372
<b>Qui</b>	1061	755	1.423186	34	0.675	1.164937	0.4960817
<b>Sex</b>	1652	1896	2.2954	47	0.4608149	9.206163	0.4493391
<b>Sab</b>	2289	3051	2.665793	46	0.5085965	6.494987	0.2985928

**Tabela B.22:** Redes SiST sociais de GOW\_SCE para  $\Delta t = 30$  min.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	2786	5666	4.06748	46	0.4431976	5.327428	0.1974859
<b>Seg</b>	2133	4522	4.240038	58	0.4127144	5.244146	0.2361808
<b>Ter</b>	1975	3808	3.856203	28	0.4147579	5.551932	0.2358822
<b>Qua</b>	1563	1739	2.225208	20	0.5401404	5.441452	0.4834692
<b>Qui</b>	1161	855	1.472868	34	0.6678395	1.25	0.4207999
<b>Sex</b>	1844	2511	2.723427	42	0.4142295	7.354545	0.3367472
<b>Sab</b>	2454	3774	3.075795	46	0.4694547	5.766256	0.252374

**Tabela B.23:** Redes SiST sociais de GOW\_SCE para  $\Delta t = 1$  hora.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	1125	1154	2.051556	18	0.5341434	4.09301	0.3538221
<b>Seg</b>	2249	5387	4.790574	32	0.3994987	5.037393	0.1887488
<b>Ter</b>	2117	4709	4.448748	31	0.3999222	5.212998	0.1876287
<b>Qua</b>	1681	2098	2.496133	20	0.5261333	5.282897	0.4888734
<b>Qui</b>	1280	961	1.501562	34	0.6530822	1.328811	0.3947134
<b>Sex</b>	2015	3271	3.24665	44	0.4101037	6.162485	0.1480987
<b>Sab</b>	2632	4465	3.392857	53	0.4487074	5.484606	0.1988944

**Tabela B.24:** Redes SiST sociais de GOW\_SCE para  $\Delta t = 2$  horas.

Rede	nós	conexões	grau	diâmetro	coef. agrup.	distância	assortatividade
<b>Dom</b>	102	54	1.058824	2	0.5	1.052632	0.4807692
<b>Seg</b>	30	17	1.133333	4	0.7777778	1.105263	0.4333333
<b>Ter</b>	20	10	1	2	NaN	1	NaN
<b>Qua</b>	27	20	1.481481	4	0.5833333	1.366667	0.4413674
<b>Qui</b>	47	30	1.276596	8	0.4333333	1.317073	0.1663061
<b>Sex</b>	64	39	1.21875	3	0.7878788	1.136364	0.5282258
<b>Sab</b>	84	46	1.095238	2	0.75	1.041667	0.452381

**Tabela B.25:** Redes SiST não sociais de TWT\_CHE para  $\Delta t = 15$  min.

<b>Rede</b>	<b>nós</b>	<b>conexões</b>	<b>grau</b>	<b>diâmetro</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>assortatividade</b>
<b>Dom</b>	131	76	1.160305	3	0.8039216	1.119048	0.4242424
<b>Seg</b>	43	27	1.255814	4	0.5833333	1.1	0.5120482
<b>Ter</b>	44	27	1.227273	3	0.3333333	1.285714	0.2801397
<b>Qua</b>	45	32	1.422222	7	0.7041667	1.347826	0.4870882
<b>Qui</b>	60	41	1.366667	6	0.6944444	1.235294	0.2746835
<b>Sex</b>	100	65	1.3	3	0.7222222	1.171053	0.6195515
<b>Sab</b>	125	75	1.2	2	0.7971014	1.0625	0.4730679

**Tabela B.26:** Redes SiST não sociais de TWT\_CHE para  $\Delta t = 30$  min.

<b>Rede</b>	<b>nós</b>	<b>conexões</b>	<b>grau</b>	<b>diâmetro</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>assortatividade</b>
<b>Dom</b>	171	114	1.333333	4	0.8266667	1.161538	0.666413
<b>Seg</b>	50	36	1.44	4	0.8333333	1.076923	0.321908
<b>Ter</b>	54	44	1.62963	6	0.6666667	1.558824	0.4193548
<b>Qua</b>	77	57	1.480519	6	0.6377778	1.256757	0.6012599
<b>Qui</b>	81	66	1.62963	7	0.7333333	1.326087	0.4208645
<b>Sex</b>	142	109	1.535211	3	0.8760684	1.091667	0.5642249
<b>Sab</b>	200	124	1.24	3	0.8536585	1.096296	0.4284165

**Tabela B.27:** Redes SiST não sociais de TWT\_CHE para  $\Delta t = 1$  hora.

<b>Rede</b>	<b>nós</b>	<b>conexões</b>	<b>grau</b>	<b>diâmetro</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>assortatividade</b>
<b>Dom</b>	237	181	1.527426	3	0.9050926	1.062176	0.6117102
<b>Seg</b>	78	56	1.435897	8	0.8	1.138462	0.3959314
<b>Ter</b>	65	61	1.876923	7	0.8868132	1.253165	0.6481346
<b>Qua</b>	92	75	1.630435	10	0.836	1.407767	0.598905
<b>Qui</b>	118	99	1.677966	7	0.7408602	1.196721	0.6562259
<b>Sex</b>	175	146	1.668571	3	0.9124294	1.058065	0.5003684
<b>Sab</b>	286	194	1.356643	3	0.84	1.115741	0.4591219

**Tabela B.28:** Redes SiST não sociais de TWT\_CHE para  $\Delta t = 2$  horas.

## C

**Propriedades das Redes Aleatórias**

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.01789157	4.953505	11
<b>Seg</b>	0.01884666	4.723251	11
<b>Ter</b>	0.02041361	5.553146	13
<b>Qua</b>	0.01527166	10.38254	31
<b>Qui</b>	0.02839506	4.071813	14
<b>Sex</b>	0.01593933	5.639454	13
<b>Sab</b>	0.01257381	5.602156	14

**Tabela C.1:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.2.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.01645655	4.261401	9
<b>Seg</b>	0.02547334	4.045077	8
<b>Ter</b>	0.02124089	4.80277	12
<b>Qua</b>	0.01271033	7.527774	20
<b>Qui</b>	0	7.921551	25
<b>Sex</b>	0.01534933	4.774858	11
<b>Sab</b>	0.01352541	4.546981	10

**Tabela C.2:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.3.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.02129483	3.700444	7
<b>Seg</b>	0.02490976	3.706226	7
<b>Ter</b>	0.02174928	4.119	9
<b>Qua</b>	0.01219489	6.21943	15
<b>Qui</b>	0.01741638	9.710667	24
<b>Sex</b>	0.01574125	4.141961	9
<b>Sab</b>	0.01452352	10.7153	32

**Tabela C.3:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.4.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.02700348	4.29	12
<b>Seg</b>	0.004605777	4.758728	12
<b>Ter</b>	0.03169643	2.971714	8
<b>Qua</b>	0.01764706	2.748768	9
<b>Qui</b>	0.07738095	2.408	7
<b>Sex</b>	0.05277778	2.551402	8
<b>Sab</b>	0	5.375102	16

**Tabela C.4:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.5.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.01141975	6.209486	17
<b>Seg</b>	0.03054275	7.499201	19
<b>Ter</b>	0.0173731	7.448443	18
<b>Qua</b>	0	3.754505	10
<b>Qui</b>	0	1.452381	3
<b>Sex</b>	0.01833333	6.53347	19
<b>Sab</b>	0.01111111	4.212541	12

**Tabela C.5:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.6.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.01601784	8.013592	22
<b>Seg</b>	0.01959817	7.845854	24
<b>Ter</b>	0.03053467	8.29472	23
<b>Qua</b>	0.0122549	3.093946	9
<b>Qui</b>	0	2.853081	8
<b>Sex</b>	0.01194196	7.046251	17
<b>Sab</b>	0.003829504	10.14224	24

**Tabela C.6:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.7.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.01890673	7.714917	21
<b>Seg</b>	0.03221745	9.209151	29
<b>Ter</b>	0.01598174	8.867543	25
<b>Qua</b>	0.0304878	2.787698	7
<b>Qui</b>	0.03191489	2.769517	7
<b>Sex</b>	0.02104709	10.23074	28
<b>Sab</b>	0.02205128	5.950717	15

**Tabela C.7:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.8.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.02427001	2.748943	4
<b>Seg</b>	0.03299383	2.620206	4
<b>Ter</b>	0.02389299	2.796688	4
<b>Qua</b>	0.009311341	4.931164	10
<b>Qui</b>	0.0008438819	16.3416	50
<b>Sex</b>	0.0235239	2.891801	4
<b>Sab</b>	0.02445598	2.747758	4

**Tabela C.8:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.9.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.03611816	2.495652	4
<b>Seg</b>	0.04934476	2.316702	3
<b>Ter</b>	0.03711825	2.536592	4
<b>Qua</b>	0.01126687	3.896206	7
<b>Qui</b>	0.003146183	11.1546	34
<b>Sex</b>	0.03626354	2.58414	4
<b>Sab</b>	0.03824228	2.473401	4

**Tabela C.9:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.10.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.04985711	2.232067	3
<b>Seg</b>	0.0645147	2.102361	3
<b>Ter</b>	0.05081331	2.278498	3
<b>Qua</b>	0.0155098	3.278958	6
<b>Qui</b>	0.003168878	7.442196	19
<b>Sex</b>	0.05678387	2.233	3
<b>Sab</b>	0.05414254	2.184305	3

**Tabela C.10:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.11.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.02182445	3.16416	5
<b>Seg</b>	0.075654	2.011822	3
<b>Ter</b>	0.06418719	2.097035	3
<b>Qua</b>	0.02077153	2.888734	4
<b>Qui</b>	0.007034314	5.688424	13
<b>Sex</b>	0.08340939	1.999979	3
<b>Sab</b>	0.07383778	2.004698	3

**Tabela C.11:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.12.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.004903455	16.24885	44
<b>Seg</b>	0.0004577062	12.37945	34
<b>Ter</b>	0	7.508266	21
<b>Qua</b>	0	3.039919	11
<b>Qui</b>	0.005555556	3.277702	13
<b>Sex</b>	0	7.364439	27
<b>Sab</b>	0.002028509	13.77625	39

**Tabela C.12:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.13.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.0003912363	13.94351	43
<b>Seg</b>	0.00412721	12.14485	32
<b>Ter</b>	0.001173223	11.96209	41
<b>Qua</b>	0	5.247339	17
<b>Qui</b>	0	3.554502	13
<b>Sex</b>	0.001067169	11.96343	45
<b>Sab</b>	0.004249359	11.48824	30

**Tabela C.13:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.14.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.002286255	10.83341	27
<b>Seg</b>	0.001965966	10.37044	30
<b>Ter</b>	0.001731199	13.42727	39
<b>Qua</b>	0.00513468	4.812854	16
<b>Qui</b>	0	2.517407	10
<b>Sex</b>	0.004017583	13.25599	35
<b>Sab</b>	0.002777303	9.614348	23

**Tabela C.14:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.15.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.001033835	7.923936	22
<b>Seg</b>	0.004688316	8.713563	25
<b>Ter</b>	0.004411711	10.00375	27
<b>Qua</b>	0.009442928	7.154057	24
<b>Qui</b>	0.004145937	3.051563	11
<b>Sex</b>	0.008092958	10.2893	30
<b>Sab</b>	0.002839814	7.871822	24

**Tabela C.15:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.16.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.01266803	2.830022	4
<b>Seg</b>	0.018364	2.712756	4
<b>Ter</b>	0.01446617	2.831003	4
<b>Qua</b>	0.003659329	5.203217	11
<b>Qui</b>	0.0009122983	12.13723	32
<b>Sex</b>	0.008700608	3.301954	5
<b>Sab</b>	0.007535457	3.322488	5

**Tabela C.16:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.17.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.01654742	2.696116	4
<b>Seg</b>	0.02625776	2.469483	3
<b>Ter</b>	0.02081358	2.633546	4
<b>Qua</b>	0.004414679	4.085628	8
<b>Qui</b>	0.002120384	7.550791	19
<b>Sex</b>	0.01394258	2.835014	4
<b>Sab</b>	0.01059492	2.899187	4

**Tabela C.17:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.18.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.02134746	2.523682	3
<b>Seg</b>	0.03341196	2.26539	3
<b>Ter</b>	0.02851409	2.392829	3
<b>Qua</b>	0.006868132	3.446109	6
<b>Qui</b>	0.002581197	5.494268	11
<b>Sex</b>	0.02213192	2.594978	4
<b>Sab</b>	0.01510898	2.728506	4

**Tabela C.18:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.19.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.01407104	2.893388	4
<b>Seg</b>	0.03786324	2.153176	3
<b>Ter</b>	0.03577719	2.199312	3
<b>Qua</b>	0.01023778	2.969395	4
<b>Qui</b>	0.00384794	4.380717	8
<b>Sex</b>	0.0328841	2.282092	3
<b>Sab</b>	0.01959197	2.561769	3

**Tabela C.19:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.20.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.001513838	14.47626	39
<b>Seg</b>	0.002953394	13.67915	40
<b>Ter</b>	0.001092821	16.21051	41
<b>Qua</b>	0.0008521303	5.197433	23
<b>Qui</b>	0	2.84619	11
<b>Sex</b>	0.0003683241	12.39285	37
<b>Sab</b>	0.001453287	17.69216	48

**Tabela C.20:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.21.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.0005963749	13.32104	36
<b>Seg</b>	0.001372901	11.18108	29
<b>Ter</b>	0.00112726	13.00029	33
<b>Qua</b>	0.001363875	16.22909	51
<b>Qui</b>	0	2.697494	12
<b>Sex</b>	0.003313188	22.52695	69
<b>Sab</b>	0.001483033	17.54687	48

**Tabela C.21:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.22.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.002351816	10.10737	25
<b>Seg</b>	0.001743651	9.488757	23
<b>Ter</b>	0.002077906	10.32459	24
<b>Qua</b>	0	17.94203	52
<b>Qui</b>	0	3.323995	15
<b>Sex</b>	0.001120549	16.81915	41
<b>Sab</b>	0.001150213	14.51922	38

**Tabela C.22:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.23.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.002456647	5.85095	19
<b>Seg</b>	0.001949267	8.390069	20
<b>Ter</b>	0.002999382	9.085645	23
<b>Qua</b>	0	21.31637	57
<b>Qui</b>	0.004036909	3.798936	17
<b>Sex</b>	0.001692282	13.22605	36
<b>Sab</b>	0.001988436	12.61647	32

**Tabela C.23:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.24.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0	1.6	4
<b>Seg</b>	0	1.647059	3
<b>Ter</b>	0.3	1.826087	4
<b>Qua</b>	0.08333333	1.916667	4
<b>Qui</b>	0.05128205	1.589286	3
<b>Sex</b>	0	1.492308	3
<b>Sab</b>	0	1.6125	5

**Tabela C.24:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.25.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0	1.495798	4
<b>Seg</b>	0.05208333	1.568182	4
<b>Ter</b>	0.1071429	2.239437	5
<b>Qua</b>	0.1280702	2.395833	7
<b>Qui</b>	0	2.121212	6
<b>Sex</b>	0	2.27381	6
<b>Sab</b>	0	2.018634	6

**Tabela C.25:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.26.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0.02205882	2.711864	7
<b>Seg</b>	0.04545455	1.763158	5
<b>Ter</b>	0.0625	2.632258	7
<b>Qua</b>	0.02380952	2.900826	8
<b>Qui</b>	0	5.265795	16
<b>Sex</b>	0	3.666667	13
<b>Sab</b>	0	1.976744	6

**Tabela C.26:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.27.

<b>Rede</b>	<b>coef. agrup.</b>	<b>distância</b>	<b>diâmetro</b>
<b>Dom</b>	0	2.535587	7
<b>Seg</b>	0	2.47191	7
<b>Ter</b>	0.02083333	2.61809	7
<b>Qua</b>	0	4.571078	14
<b>Qui</b>	0.04269006	3.419598	10
<b>Sex</b>	0	2.933054	10
<b>Sab</b>	0.01315789	2.29918	8

**Tabela C.27:** Propriedades das redes aleatórias de mesmo tamanho das redes da Tabela B.28.

## D

## Tabelas Adicionais do Teste de Vuong do Capítulo 5

<b>Rede</b>	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.47	-0.7435025	0.4571775	Power-law
Seg	0.69	-0.4667896	0.6406505	Power-law
Ter	0.39	-0.08038102	0.9359342	Power-law
Qua	0.84	-0.9129763	0.361255	Power-law
Qui	0.8	-0.3460883	0.7292763	Power-law
Sex	0.47	-0.9419218	0.3462327	Power-law
Sab	0.47	-0.4352341	0.6633925	Power-law

Tabela D.1: Teste de Vuong para redes SiST não sociais da base BKT\_CHO para  $\Delta t = 15$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.31	-1.144794	0.2522945	Power-law
Seg	0.21	-0.9930682	0.3206767	Power-law
Ter	0.31	-1.03361	0.3013185	Power-law
Qua	0.09	-1.844608	0.06509455	Inconclusivo
Qui	0.36	-0.3130083	0.7542744	Power-law
Sex	0.03	-2.296897	0.02162462	Inconclusivo
Sab	0.42	-0.9487093	0.3427685	Power-law

Tabela D.2: Teste de Vuong para redes SiST não sociais da base BKT\_CHO para  $\Delta t = 30$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.31	-1.144794	0.2522945	Power-law
Dom	0.11	-1.157185	0.2471969	Power-law
Seg	0.86	-0.3048225	0.7605013	Power-law
Ter	0.35	-0.3564686	0.7214897	Power-law
Qua	0.7	-0.4801748	0.6311031	Power-law
Qui	0.27	-0.6221045	0.5338731	Power-law
Sex	0.75	-0.1388741	0.8895496	Power-law
Sab	0	-2.800643	0.005100089	Inconclusivo

Tabela D.3: Teste de Vuong para redes SiST não sociais da base BKT\_CHO para  $\Delta t = 1$  hora.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$P_{pl-ln}$	<b>Resultado</b>
Dom	0.31	-1.144794	0.2522945	Power-law
Dom	0.37	-1.163981	0.2444318	Power-law
Seg	0.98	-0.258745	0.795832	Power-law
Ter	0	-6.477284	9.34E-011	Inconclusivo
Qua	0.85	-0.4719032	0.6369959	Power-law
Qui	0.53	-0.5174959	0.60481	Power-law
Sex	0.99	-0.4545015	0.6494679	Power-law
Sab	0.32	-0.7380043	0.4605118	Power-law

**Tabela D.4:** Teste de Vuong para redes SiST não sociais da base BKT\_CHO para  $\Delta t = 2$  horas.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$P_{pl-ln}$	<b>Resultado</b>
Dom	0.31	-1.144794	0.2522945	Power-law
Dom	0.61	-0.9328139	0.3509161	Power-law
Seg	0.62	-0.9854815	0.3243876	Power-law
Ter	0.77	-0.6696672	0.50307	Power-law
Qua	0.02	-0.8253689	0.4091622	Inconclusivo
Qui	0.02	-0.6006081	0.548101	Inconclusivo
Sex	0.04	-1.138051	0.2550993	Inconclusivo
Sab	0.53	-0.2339516	0.8150225	Power-law

**Tabela D.5:** Teste de Vuong para redes SiST sociais da base BKT\_CHO para  $\Delta t = 15$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$P_{pl-ln}$	<b>Resultado</b>
Dom	0.31	-1.144794	0.2522945	Power-law
Dom	0.38	-0.6312343	0.5278873	Power-law
Seg	0.36	-1.28416	0.1990859	Power-law
Ter	0.46	-0.651372	0.5148064	Power-law
Qua	0	-0.9209699	0.3570662	Inconclusivo
Qui	0.91	-0.4259474	0.6701461	Power-law
Sex	0.75	-0.6820806	0.495188	Power-law
Sab	0.02	-1.171434	0.2414243	Inconclusivo

**Tabela D.6:** Teste de Vuong para redes SiST sociais da base BKT\_CHO para  $\Delta t = 30$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$P_{pl-ln}$	<b>Resultado</b>
Dom	0.31	-1.144794	0.2522945	Power-law
Dom	0.03	-2.104687	0.03531854	Inconclusivo
Seg	0.45	-1.294745	0.1954083	Power-law
Ter	0.84	-0.6124346	0.5402503	Power-law
Qua	0.01	-1.155729	0.2477921	Inconclusivo
Qui	0.01	-0.7291581	0.465905	Inconclusivo
Sex	0.24	-0.8267454	0.4083814	Power-law
Sab	0.64	-0.6399479	0.5222065	Power-law

**Tabela D.7:** Teste de Vuong para redes SiST sociais da base BKT\_CHO para  $\Delta t = 1$  hora.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.31	-1.144794	0.2522945	Power-law
Dom	0.01	-2.228741	0.02583112	Inconclusivo
Seg	0.34	-1.401444	0.1610813	Power-law
Ter	0.6	-0.7703732	0.4410785	Power-law
Qua	0.01	-1.278207	0.2011765	Inconclusivo
Qui	0.74	-0.4085043	0.6829035	Power-law
Sex	0.47	-0.9094684	0.363103	Power-law
Sab	0.22	-0.3022143	0.7624887	Power-law

**Tabela D.8:** Teste de Vuong para redes SiST sociais da base BKT\_CHO para  $\Delta t = 2$  horas.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.05	0.574561	-0.561347	Inconclusivo
Seg	0.14	0.2816244	-1.076677	Power-law
Ter	0	3.15E-006	-4.660586	Log-normal
Qua	0	1.67E-005	-4.304865	Log-normal
Qui	0	0.01197496	-2.512881	Log-normal
Sex	0.49	0.8751249	-0.1571522	Power-law
Sab	0.62	0.6789008	-0.4139635	Power-law

**Tabela D.9:** Teste de Vuong para redes SiST não sociais da base GOW\_CHO para  $\Delta t = 15$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0	0.1338073	-1.499256	Inconclusivo
Seg	0.07	0.1878072	-1.317094	Inconclusivo
Ter	0.32	0.2071602	-1.261413	Power-law
Qua	0	1.62E-006	-4.795464	Log-normal
Qui	0.98	0.9230948	-0.09653615	Power-law
Sex	0.16	0.5519541	-0.5948346	Power-law
Sab	0.01	0.1520533	-1.432316	Inconclusivo

**Tabela D.10:** Teste de Vuong para redes SiST não sociais da base GOW\_CHO para  $\Delta t = 30$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.18	0.3567621	-0.9215524	Power-law
Seg	0	0.007315453	-2.682132	Log-normal
Ter	0.32	0.3531152	-0.9285639	Power-law
Qua	0	2.18E-007	-5.183673	Log-normal
Qui	0	4.48E-007	-5.047229	Log-normal
Sex	0.38	0.8282479	0.2169492	Power-law
Sab	0.37	0.4661791	-0.72871	Power-law

**Tabela D.11:** Teste de Vuong para redes SiST não sociais da base GOW\_CHO para  $\Delta t = 1$  hora.

<b>Rede</b>	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.55	0.7540373	0.3133203	Power-law
Seg	0	0.002647821	-3.005916	Log-normal
Ter	0	3.39E-022	-9.687978	Log-normal
Qua	0	1.92E-008	-5.618929	Log-normal
Qui	0	3.69E-007	-5.084538	Log-normal
Sex	0.13	0.1742712	1.358607	Power-law
Sab	0	1.75E-014	-7.6681	Log-normal

**Tabela D.12:** Teste de Vuong para redes SiST não sociais da base GOW\_CHO para  $\Delta t = 2$  horas.

<b>Rede</b>	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	1	0.8736949	-0.1589669	Power-law
Seg	0.05	0.1248333	-1.534799	Inconclusivo
Ter	0	2.76E-005	-4.19239	Log-normal
Qua	0.17	0.1640617	-1.39154	Power-law
Qui	0	0.06079243	-1.875002	Log-normal
Sex	0.61	0.799084	-0.2545328	Power-law
Sab	0.03	0.07393765	-1.786999	Log-normal

**Tabela D.13:** Teste de Vuong para redes SiST sociais da base GOW\_CHO para  $\Delta t = 15$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.73	0.6332922	-0.4770982	Power-law
Seg	0.61	0.3147624	-1.005279	Power-law
Ter	0.53	0.534252	-0.6215284	Power-law
Qua	0.59	0.4925187	-0.6863083	Power-law
Qui	0.11	0.08988473	-1.696006	Power-law
Sex	0.28	0.8227446	-0.2240164	Power-law
Sab	0.53	0.9392049	-0.0762692	Power-law

**Tabela D.14:** Teste de Vuong para redes SiST sociais da base GOW\_CHO para  $\Delta t = 30$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.93	0.6429729	-0.4635462	Power-law
Seg	0.02	0.02565491	-2.231396	Log-normal
Ter	0.76	0.7803552	-0.2788561	Power-law
Qua	0.51	0.4837701	-0.7002518	Power-law
Qui	0.27	0.649055	-0.4550755	Power-law
Sex	0	0.0008011775	-3.352388	Log-normal
Sab	0	0.04758832	-1.981026	Log-normal

**Tabela D.15:** Teste de Vuong para redes SiST sociais da base GOW\_CHO para  $\Delta t = 1$  hora.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.6	0.7384161	-0.3339516	Power-law
Seg	0.07	0.06915462	-1.817408	Log-normal
Ter	0.09	0.8605239	-0.1757073	Inconclusivo
Qua	0.83	0.4112202	-0.8217482	Power-law
Qui	0.9	0.5678666	-0.5711963	Power-law
Sex	0.54	0.7806488	0.2784737	Power-law
Sab	0.09	0.04524187	-2.002399	Log-normal

**Tabela D.16:** Teste de Vuong para redes SiST sociais da base GOW\_CHO para  $\Delta t = 2$  horas.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.43	4.192581	2.76E-005	Power-law
Seg	0.75	5.296989	1.18E-007	Power-law
Ter	0	13.43015	0	Inconclusivo
Qua	0.72	4.267838	1.97E-005	Power-law
Qui	0.87	1.300058	0.1935809	Power-law
Sex	0.24	4.386435	1.15E-005	Power-law
Sab	1	4.345567	1.39E-005	Power-law

**Tabela D.17:** Teste de Vuong para redes SiST não sociais da base GOW\_SCE para  $\Delta t = 15$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.03	4.756243	1.97E-006	Inconclusivo
Seg	0.24	5.413793	6.17E-008	Power-law
Ter	0.28	1.959463	0.05005859	Power-law
Qua	0.4	3.918669	8.90E-005	Power-law
Qui	0.05	3.006363	0.002643933	Inconclusivo
Sex	0.45	4.498723	6.84E-006	Power-law
Sab	0.78	4.90271	9.45E-007	Power-law

**Tabela D.18:** Teste de Vuong para redes SiST não sociais da base GOW\_SCE para  $\Delta t = 30$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.51	5.96338	2.47E-009	Power-law
Seg	0	11.77875	0	Inconclusivo
Ter	0	17.0892	0	Inconclusivo
Qua	0.1	3.779971	0.0001568465	Inconclusivo
Qui	0.01	5.869442	4.37E-009	Inconclusivo
Sex	0.17	4.826563	1.39E-006	Power-law
Sab	0.43	5.251118	1.51E-007	Power-law

**Tabela D.19:** Teste de Vuong para redes SiST não sociais da base GOW\_SCE para  $\Delta t = 1$  hora.

<b>Rede</b>	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.01	7.094637	1.30E-012	Inconclusivo
Seg	0.37	6.412933	1.43E-010	Power-law
Ter	0.16	3.887569	0.0001012532	Power-law
Qua	0.09	2.789673	0.005276136	Inconclusivo
Qui	0	2.743879	0.006071787	Inconclusivo
Sex	34	6.678981	2.41E-011	Power-law
Sab	0	9.012553	0	Inconclusivo

**Tabela D.20:** Teste de Vuong para redes SiST não sociais da base GOW\_SCE para  $\Delta t = 2$  horas.

<b>Rede</b>	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0	6.318894	2.63E-010	Inconclusivo
Seg	0.05	4.597975	4.27E-006	Inconclusivo
Ter	0.01	3.638644	0.0002740772	Inconclusivo
Qua	0.27	1.663435	0.09622533	Power-law
Qui	0	3.543015	0.0003955806	Inconclusivo
Sex	0.55	1.969537	0.04889144	Power-law
Sab	0.2	2.756651	0.005839661	Power-law

**Tabela D.21:** Teste de Vuong para redes SiST sociais da base GOW\_SCE para  $\Delta t = 15$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.43	2.625669	0.008647896	Power-law
Seg	0.03	4.263006	2.02E-005	Inconclusivo
Ter	0.09	4.76585	1.88E-006	Inconclusivo
Qua	0.1	1.179141	0.2383421	Inconclusivo
Qui	0	3.647336	0.000264973	Inconclusivo
Sex	0	7.46912	8.08E-014	Inconclusivo
Sab	0.73	2.068844	0.0385607	Power-law

**Tabela D.22:** Teste de Vuong para redes SiST sociais da base GOW\_SCE para  $\Delta t = 30$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.73	2.519231	0.01176116	Power-law
Seg	0.01	4.285824	1.82E-005	Inconclusivo
Ter	0.56	3.158847	0.001583948	Power-law
Qua	0.19	1.708735	0.08750005	Power-law
Qui	0	4.686783	2.78E-006	Inconclusivo
Sex	0	8.26818	2.22E-016	Inconclusivo
Sab	0.94	2.644658	0.008177345	Power-law

**Tabela D.23:** Teste de Vuong para redes SiST sociais da base GOW\_SCE para  $\Delta t = 1$  hora.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.27	2.507032	0.01217495	Power-law
Seg	0.34	2.988698	2.80E-003	Power-law
Ter	0.04	6.295005	3.07E-010	Inconclusivo
Qua	0.07	3.079269	0.00207509	Inconclusivo
Qui	0.05	4.491098	7.09E-006	Inconclusivo
Sex	0	8.415092	0	Inconclusivo
Sab	0.27	5.678831	1.36E-008	Power-law

**Tabela D.24:** Teste de Vuong para redes SiST sociais da base GOW\_SCE para  $\Delta t = 2$  horas.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.27	-4.007882	6.13E-005	Power-law
Seg	0.27	0.3791494	7.05E-001	Power-law
Ter	-	-	-	Power-law
Qua	0.27	1.58601	1.13E-001	Power-law
Qui	0.27	1.508982	0.1313034	Power-law
Sex	0.27	-0.718098	4.73E-001	Power-law
Sab	0.27	-7.943681	1.96E-015	Power-law

**Tabela D.25:** Teste de Vuong para redes SiST não sociais da base TWT\_CHE para  $\Delta t = 15$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.02	-1.839681	6.58E-002	Inconclusivo
Seg	0.02	1.767097	7.72E-002	Inconclusivo
Ter	0.02	-0.787364	4.31E-001	Inconclusivo
Qua	0.02	1.898991	0.05756571	Inconclusivo
Qui	0.86	1.187396	0.2350713	Power-law
Sex	0.21	0.603356	5.46E-001	Power-law
Sab	0	-2.58269	0.009803331	Inconclusivo

**Tabela D.26:** Teste de Vuong para redes SiST não sociais da base TWT\_CHE para  $\Delta t = 30$  min.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.02	0.1357382	8.92E-001	Inconclusivo
Seg	0.05	0.5656043	0.5716628	Inconclusivo
Ter	0.06	-0.65709	5.11E-001	Inconclusivo
Qua	0.08	1.976478	4.81E-002	Inconclusivo
Qui	0.22	2.028879	0.04247059	Power-law
Sex	0.06	1.037786	0.2993698	Inconclusivo
Sab	0	-2.856562	4.28E-003	Inconclusivo

**Tabela D.27:** Teste de Vuong para redes SiST não sociais da base TWT\_CHE para  $\Delta t = 1$  hora.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
Dom	0.18	-0.2669469	0.78951	Power-law
Seg	0.59	0.1041284	0.9170674	Power-law
Ter	0.01	1.016436	3.09E-001	Inconclusivo
Qua	0.83	1.822889	6.83E-002	Power-law
Qui	0.47	2.66026	0.007808045	Power-law
Sex	0	-0.4581636	0.6468349	Inconclusivo
Sab	0.57	0.02291375	0.9817191	Power-law

**Tabela D.28:** Teste de Vuong para redes SiST não sociais da base TWT\_CHE para  $\Delta t = 2$  horas.

<b>Rede</b>	$p_{pl}$	$\mathcal{LR}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
1	0.77	-0.2796212	7.80E-001	Power-law
2	0.01	1.162011	0.245231	Inconclusivo
3	0	4.099489	4.14E-005	Inconclusivo
4	0.11	-1.847852	6.46E-002	Power-law
5	0.94	0.8471784	0.3968957	Power-law
6	0.1	-0.7600244	0.44724	Inconclusivo
7	0.05	-1.360168	1.74E-001	Inconclusivo

**Tabela D.29:** Teste de Vuong para redes SiST não sociais da base 4QR\_yua para  $\Delta t = 2$  horas.

<b>Rede</b>	$p_{pl}$	$\mathcal{L}\mathcal{R}_{pl-ln}$	$p_{pl-ln}$	<b>Resultado</b>
1	0.07	2.105674	0.03523271	Inconclusivo
2	0.64	5.307595	1.11E-007	Power-law
3	0	15.81024	0	Inconclusivo
4	0.07	3.615228	0.0003000829	Inconclusivo
5	0	9.55329	0.00E+000	Inconclusivo
6	0.01	2.919972	3.50E-003	Inconclusivo

**Tabela D.30:** Teste de Vuong para redes SiST não sociais da base TWT\_EIN para  $\Delta t = 2$  horas.