

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

RENATO FABBRI

Redes complexas para o participante

São Carlos

2015

RENATO FABBRI

Complex networks for the participant

Monografia apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para o Exame de Qualificação como parte dos requisitos para obtenção do título de Doutor em Ciências.

Área de concentração: Física Básica
Orientador: Prof. Dr. Osvaldo Novais de Oliveira Jr.

São Carlos

2015

RESUMO

FABBRI, C. *Redes complexas para o participante*. Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2015.

As redes complexas formam uma das áreas mais ativas da física redente. Há esforços consideráveis para apresentar estes avanços ao público geral, mas tudo indica que poucos ou nenhum são voltados para o benefício do indivíduo que constitui estes sistemas. Ou seja, com um núcleo de conhecimento da área, e receitas para aproveitamento, fornece meios para o participante interagir e entender as redes nas quais ele se encontra. Este trabalho objetiva realizar tal tarefa por meio das redes sociais do participante. Verificamos que tais redes exibem uma estabilidade temporal de medidas temporais e dos tamanhos relativos dos setores conectivos básicos (hubs, intermediários, periféricos). Observamos uma diferenciação da produção de texto de cada setor básico. Também formalizamos as conceitualizações destas redes como OWL onde foi possível, principalmente as instâncias de participação dosial previstas por lei e implementadas como software. Software e dados foram disponibilizados e usados. Protocolos escolhidos para facilitar a integração de estruturas de diferentes procedências, para reutilização dos dados em outros trabalhos e pesquisas, e para o benefício público. Consequências conceituais requerem considerações antropológicas e estão sendo documentadas. Próximos passos são: melhor documentação e desenvolvimento do aparato em software, ontologias e dados; considerações tipológicas das propriedades físicas observadas nas redes de interação humana, com atenção aos outliers, às relações entre topologia do agente e texto produzido, e à ponte com a bagagem mais tradicional das ciências humanas no assunto.

Palavras-chave: Redes complexas. Redes sociais. Complexidade. Física antropológica. Dados ligados. Web semântica. Participação social. Mineração de texto. Processamento de linguagem natural.

Sumário

1	Introdução	7
1.1	Revisão de literatura	8
1.1.1	Ambiguidades e sinonimos no jargão	9
1.1.2	Processamento de linguagem natural, dados ligados, participação social	10
2	Materiais	11
2.1	O banco de dados públicos sobre listas de email (benchmark)	11
2.2	Facebook, Twitter, Participa.br, Cidade Democrática, AA	11
2.3	Minhas próprias redes sociais	11
3	Methods	13
3.1	Circular statistics	14
3.2	Erdős Sectioning	14
3.3	PCA of measures along time	14
3.4	Kolmogorov-Smirnoff test for texts produced by sectors	14
3.5	Audiovisualization of data	14
3.6	Typological considerations	14
3.7	Semantic web	14
3.7.1	OWL ontology construction	14
4	Results	15
4.1	Time stability in human interaction networks	16
4.2	Semantic web	16
4.2.1	Linked data	16
4.2.2	RDF data conversion of data into linked data	16
4.2.3	Published linked data and OWL ontologies	16
4.3	Harnessing	16
4.3.1	Social percolation procedures	16

4.3.2	Recommendation systems for the enrichment of semantic navigation . .	16
4.3.3	Understanding the social being	16
5	Finished and planed tasks, chronogram	17
5.1	Documents	17
5.1.1	To be finished	17
5.1.2	Finished	18
5.2	Cronograma	18
6	Conclusions	21

1

Introdução

Estudos sobre redes de interação humana foram iniciados bem antes dos computadores modernos, datam do século XIX, enquanto a fundação da “Análise de Redes Sociais”/ARS (*Social Network Analysis*/SNA) é geralmente atribuída ao psiquiatra Jacob Moreno na metade do século vinte (?). Com a crescente disponibilidade de dados relacionados à interação humana, a pesquisa destas redes tem aumentado continuamente. Contribuições podem ser encontradas em uma variedade de áreas, de ciências sociais e humanidades (?) a ciências sociais (?) e física (?, ?), dada a natureza multidisciplinar do assunto. Uma das abordagens da perspectiva de uma ciência exata é representar a rede de interação como uma rede complexa (?, ?), com a qual algumas características foram reveladas. Por exemplo, a topologia das redes de interação humana exibem um traço livre de escala, o que aponta para a existência de um pequeno número de hubs super conectados e um grande número de vértices pouco conectados.

Há um hiato de conhecimento e tecnologia entre o legado de redes complexas e o usufruto do participante. Este hiato é reativo, e há evidência de que conseguirá se manter como um ecossistema de conhecimento, tecnologia e empreendimento da sociedade em todas as suas escalas. Deve facilitar, por exemplo: elaboração e preparação de documentos, aquisição rápida de conhecimento, realização de empreitadas coletivas. Em geral: processos de coleta e difusão de informação (e bens).

Este trabalho apresenta uma confirmação deste cenário e avanços. Algumas estratégias foram selecionadas para verificar a aplicabilidade de conceitos de redes complexas para o benefício do participante. Em especial, experimentos muito simples parecem capazes de modificar estruturas sociais. Neste contexto, verificamos estabilidades temporais nas redes de interação humana, e expomos que os setores primitivos das redes (hubs, intermediários e periféricos) produzem textos bastante diferentes entre si. Este conhecimento é útil para uma tipologia não estigmatizante de participantes em redes de interação. A audiovisualização e interconexão de dados com arte e engenhocas em software deram suporte contínuo à pesquisa científica.

Aplicações foram complementadas com a Presidência da República e o PNUD/ONU.

A próxima seção apresenta considerações gerais sobre a literatura. A Seção 2 é dedicada aos dados analisados. A Seção ?? contém os métodos usados para atingir os resultados, que são explicitados na Seção ?. O cronograma de atividades e uma comparação entre afazeres planejados, em andamento e finalizados estão na Seção ?. A monografia termina com as conclusões na Seção ?, seguida de agradecimentos e bibliografia.

1.1 Revisão de literatura

A área das redes complexas é relativamente nova (≈ 25 anos) e a literatura apresenta definições divergentes da área em si. Uma definição que tem recebido aceitação crescente é da rede complexa como “um grafo grande com características topológicas não triviais”. Esta definição é enganosa ao menos em três pontos. Primeiro, há redes de interesse com características topológicas triviais, como as redes de Erdős-Rényi e a Geográfica (?), ou as redes simples usadas para exemplos. Segundo, a definição falha ao não emitir a mensagem fundamental de que uma rede complexa não é somente uma estrutura matemática, um grafo isolado. As redes complexas de interesse são redes reais ou modelos idealizados para as entender. Além disso, não só grafos grandes são de interesse, mas grafos pequenos são comumente usados como exemplos de propriedades e extensão das estruturas maiores. Uma definição, ainda longe de perfeita, mas preferida neste trabalho, é considerar a área das redes complexas como interessada em “redes usualmente grandes, consideradas no, ou para consideração do, meio em que residem”. This definition resolves both issues.

Os livros em geral apresentam um comum e poderoso repertório para a caracterização de sistemas complexos através de grafos. Talvez mais notáveis sejam:

- O arsenal de medidas: grau, força, betweenness centrality, coeficiente de clusterização, etc.
- Os paradigmas básicos de redes: Erdős-Rényi, geográfica, de mundo pequeno e livre de escala.
- A abordagem transdisciplinar para considerar o meio no qual a rede está inserida, ou

que implica na rede.

A literatura sobre análise de redes sociais (ARS, ou *SNA* para *Social Network Analysis*), por exemplo, pode ser frequentemente compreendida como redes complexas em sistemas sociais humanos.

Uma consideração cuidadosa dos livros e artigos lidos para esta pesquisa estão na Seção ??.

As seções a seguir (1.1.1 e 1.1.2) explicitam peculiaridades do jargão da área e considerações sobre as áreas secundárias.

1.1.1 Ambiguidades e sinonimos no jargão

A área de redes complexas é recente e conflui com diversas correntes científicas, como a física, a biologia e a sociologia. Assim, possui termos ambiguos e sinònimos.

Exemplos de ambiguidade e delimitações adotadas:

- Os vértices mais conectados são, por definição, chamados hubs da rede. O vértice mais conectado é chamado hub da rede. No contexto do algoritmo HITS, o que é bem comum, estes significados mudam: os hubs são os que possuem mais arestas saindo (grau de saída); as autoridades recebem as arestas, ou são referenciados por vários hubs e outras entidades.
- Há uma definição de centro e periferia com relação ao raio e diâmetro da rede (?, ?). Por extensão os intermediários podem ser considerados os que não são centro nem periferia. Esta setorialização centro, intermediários e periferia gera frações que diferem do previsto pela literatura para as frações de hubs, intermediários e periféricos. Um método apropriado para realizar esta setorialização da rede, com resultados estáveis e significativos, consta na Seção ??.
- etc

1.1.2 Processamento de linguagem natural, dados ligados, participação social

Os termos processamento de linguagem natural (PLN) e mineração de texto (MT) podem em geral serem substituídos um pelo outro. O termo PLN é preferido pois os intuitos da pesquisa são muito mais próximos aos intuitos da área: compreender como a linguagem verbal está sendo usada para significar.

Os termos web semântica e dados ligados em geral também podem ser substituídos um pelo outro. O primeiro salienta a rede de referenciamento dos dados, o segundo os dados referenciando-se. Principalmente na esfera acadêmica, a área é, salvo segunda ordem, sinônimo de dados em RDF via XML ou Turtle, ontologias OWL e máquinas de inferência.

A participação social é a incorporação da própria sociedade nos processos de governança da sociedade. Quase toda a participação social atual é indireta e presencial, com a população fornecendo diretrizes e indicadores para o setor público. A transparência tem sido cada vez mais presente, e o norte de “participação direta” (participação direta da sociedade civil na tomada de decisões pelo Estado) cada vez mais presente.

2

Materiais

2.1 O banco de dados públicos sobre listas de email (benchmark)

2.2 Facebook, Twitter, Participa.br, Cidade Democrática, AA

2.3 Minhas próprias redes sociais

Considerações sobre o uso de minhas anotações para pesquisa sobre mim e meu mundo.

3 Methods

3.1 Circular statistics

3.2 Erdős Sectioning

3.3 PCA of measures along time

3.4 Kolmogorov-Smirnoff test for texts produced by sectors

3.5 Audiovisualization of data

3.6 Typological considerations

3.7 Semantic web

3.7.1 OWL ontology construction

4 Results

4.1 Time stability in human interaction networks

4.2 Semantic web

4.2.1 Linked data

4.2.2 RDF data conversion of data into linked data

4.2.3 Published linked data and OWL ontologies

4.3 Harnessing

4.3.1 Social percolation procedures

4.3.2 Recommendation systems for the enrichment of semantic navigation

4.3.3 Understanding the social being

5

Finished and planed tasks, chronogram

5.1 Documents

5.1.1 To be finished

Anthropological physics

The study of human systems raises conceptual and ethical issues that require anthropological considerations. There are two immediate routes to this concepts:

- What data should or can be used?
- Can one experiment in a network of humans? In which context?

The short answer is that ethics committees and procedures are dedicated to dealing with those issues. Even so, there is a key-concept from the anthropological legacy: the study of the self as exposed to the interested culture or context. In this sense, it is reasonable (if not a suggestion) that a researcher do reflexive consideration, i.e. that he/she observe and make assumptions about its own sampling of the world. Within this same framework, many social networks (email, Facebook, Twitter, Participa.br, AA) were openly mined, with feedback to and from the studied communities. The term “anthropological physics” started being used in Brazil around 2014 and can be thought as a subfield of Social Physics.

Gradus

Uma lista detalhada de ambiguidades e sinônimos deverá completar o que está na Seção 1.1.1.

Fazer o 2⁽¹⁰⁰²⁾

Consider a idealized constitution of these networks:

- the resources of the environment are the persons, each with an amount of time available.
- The amount of resource employed by the environment to the network is constant through all connective sectors

5.1.2 Finished

5.2 Cronograma

Atividade	2013		2014		2015	
	1°	2°	1°	2°	1°	2°
1	[•]	[•]	•	•		
2	[•]	[•]	[•]	[]	[]	[]
3	[]	[•]	[•]	[•]	[•]	[]
4	[]	[•]	[•]	[•]	[•]	[•]
5					[•]	[•]
6	[•]	[•]	[•]	[•]	[•]	[•]
7	[•]	[•]	[•]	[•]	[•]	[•]

Tabela 5.1 – Cronograma de atividades ao longo dos semestres, descritas na Seção 5.2. A marcação • indica previsão feita no início do doutorado. A marcação [] se refere ao relato e previsão, agora no final do 1° semestre de 2015. As principais diferenças do previsto foram: as disciplinas foram terminadas no primeiro ano; a revisão da literatura, os acréscimos aos modelos atuais com o foco no participante da rede, e a implementação computacional, estas três atividades estão sendo realizadas constantemente e devem durar até pouco antes da entrega e defesa da tese.

Este projeto foi inicialmente dividido segundo as etapas a seguir e usadas como referência na Tabela 5.1:

1. Créditos Obrigatórios: cumprimento dos créditos obrigatórios em disciplinas, exigidas pelo programa de Doutorado do IFSC/USP.
2. Revisão da literatura.
3. Acréscimos aos modelos atuais com o foco no participante da rede.
4. Implementação computacional.
5. Escrita da tese.
6. Escrita e publicação dos resultados em artigos.
7. Trocas com pessoas externas, estabelecimento de colaborações.

Considerações sobre estes itens:

1. Foram cursadas as disciplinas de Processamento de linguagem natural, Mineração de dados, Visualização de dados e Web semântica. Dediquei um ano inteiramente às disciplinas. Estranhamente, fechei todas com B. No mestrado, fazia mais de 30 créditos na graduação, 4 disciplinas na pós, pesquisa, e fechei todas com A.
2. A literatura de para o trabalho proposto é ampla e este aprofundamento tem sido constante,
3. Os acréscimos aos modelos atuais tem tido o foco no participante da rede.
4. Há implementação computacional de provas de conceito, bibliotecas, rotinas básicas e rotinas para replicar resultados do grupo de pesquisa.
5. A escrita da tese pode tomar vários rumos: pode consistir de um conjunto de artigos ou de uma monografia final. Acho mais provável que seja um conjunto de artigos centrados no descrito na Seção 5.1.1.
6. Conseguimos finalizar um artigo (?). Há ao menos mais um em condições de publicação (?) e outro mais indiretamente relacionado sobre música (?). Além destes, há mais estes artigos no arXiv (?, ?, ?, ?), todos referentes ao trabalho do doutorado. Foram publicados em revista internacional os artigos AA e Images/Vilson, ambos sem a colaboração do orientador.

7. Parte substancial do trabalho consistiu em experimentos de coleta e difusão de informação, o que disparou reuniões, visitas e colaborações. Este processo foi iniciado logo antes do doutorado e pode ser apreciado, por exemplo, pelas visitas a São Carlos de parcerios de pesquisa, pela integração do pesquisador ao grupo de pesquisa Nexus, vinculado ao CNPq, e ao aporte do PNUD/ONU dado ao pesquisador, sobre o qual a Presidência da República se posicionou como beneficiária.

6

Conclusions
