

REDES COMPLEXAS PARA REDES SOCIAIS: INTRODUÇÃO, ASPECTOS CRÍTICOS E SOFTWARE

*Complex networks for social networks: introduction,
critical aspects and software*

*Redes complejas a las redes sociales:
introducción, aspectos críticos y software*

RESUMO Este ensaio é dedicado a expor a base da área de redes complexas e resultados de sua aplicação para a análise de redes sociais humanas. Após uma exposição teórica, são descritos a obtenção de redes complexas sociais através de diferentes critérios e alguns aspectos-chave na caracterização desses sistemas, como a estabilidade topológica e a diferenciação do texto produzido por setores distintos. São discutidos uma tipologia de participantes de uma rede social ancorada em hubs, periféricos e intermediários e o conceito de física antropológica. O texto inclui uma exposição sobre o aparato em software para a exploração das estruturas sociais representadas como redes, tanto programas escritos por nós como por outras comunidades dedicadas. Ao final, há considerações a respeito deste documento e trabalhos futuros.

PALAVRAS-CHAVE: REDES COMPLEXAS. ANÁLISE DE REDES SOCIAIS. TIPOLOGIA HUMANA. FÍSICA ANTROPOLÓGICA. SOFTWARE.

ABSTRACT This essay is dedicated to exposing the fundamentals of the complex networks field and results of its application for the analysis of human social networks. After a theoretical exposition, we described the observation of complex social networks through different criteria and some key aspects in the characterization of these systems, such as the topological stability and the differentiation of the text produced by different sectors. A typology of participants of a social network based on hubs, peripherals and intermediates and the concept of anthropological physics are discussed. The text includes an exposition of the software apparatus for the exploitation of social structures represented as networks, both programs written by us and by other dedicated communities. At the end, there are final considerations about this document and future work.

KEY-WORDS: COMPLEX NETWORKS. SOCIAL NETWORK ANALYSIS. HUMAN TYPOLOGY. ANTHROPOLOGICAL PHYSICS. SOFTWARE.

RENATO FABBRI

Instituto de Física de
São Carlos da Universidade
de São Paulo (IFSC/USP).
São Carlos/SP – Brasil

RESUMEN En este ensayo se presenta la base del área de las redes complejas y resultados de su aplicación para el análisis de redes sociales humanas. Después de una exposición teórica se describen la obtención de redes sociales complejas utilizando diferentes criterios y algunos aspectos clave de la caracterización de estos sistemas, como la estabilidad topológica y la diferenciación de texto producido por diferentes sectores. Se discute una tipología de participantes en una red social basada en hubs, los periféricos y los intermediarios y el concepto de la física antropológica. El texto incluye una exposición sobre el aparato en software para la exploración de las estructuras sociales representadas como redes, ambos programas escritos por nosotros como por otras comunidades dedicadas. Al final, hay comentarios sobre este documento y trabajo futuro.

PALAVRAS CHAVE: REDES COMPLEJAS. ANÁLISIS DE REDES SOCIALES. TIPOLOGÍA HUMANA. FÍSICA ANTROPOLÓGICA. SOFTWARE.

1. INTRODUÇÃO

Os primeiros estudos explicitamente sobre redes sociais datam do século XIX, já a fundação da área de Análise de Redes Sociais é geralmente atribuída ao psiquiatra Jacob Moreno na metade do século XX (MORENO, 1934; NEWMAN, 2010). Com o aumento da disponibilidade de dados relacionados à interações humanas, a pesquisa a respeito dessas redes tem crescido continuamente. As contribuições são agora encontradas em uma variedade de campos, de ciências sociais e humanidades (LATOUR, 2013), ciência da computação (BIRD et al., 2006) e física (VÁZQUEZ et al., 2006), dada a multidisciplinaridade do assunto. Uma das abordagens de uma perspectiva das ciências exatas é representar as redes de interação como redes complexas (BALL; NEWMAN, 2012), pela qual algumas características das redes de interação humana foram reveladas. Por exemplo, a topologia das redes de interação humana apresenta um perfil livre de escala, o que aponta para a existência de um pequeno número de hubs altamente conectados e um grande número de participantes pouco conectados. A dinâmica de redes complexas que representam a interação humana foi abordada em diversos documentos científicos (PALLA; BARABÁSI; VICSEK, 2007), mas limitadamente, dado que a pesquisa é

normalmente focada em uma métrica ou tarefa, como acessibilidade ou detecção de comunidades (TRAVENTÇOLO; COSTA, 2008).

Há muitos artigos, livros, websites e software a respeito de redes complexas e sociais e acerca da mineração de texto em mídias sociais (NEWMAN, 2010). Há menos esforços para caracterizar essas redes para além de traços gerais como o aspecto livre de escala ou para lidar com o texto produzido por redes sociais a partir da base de redes complexas. A pesquisa com relação à evolução de redes é usualmente restrita ao crescimento das redes, em que há um aumento monotônico no número de eventos (PALLA; BARABÁSI; VICSEK, 2007). Tipos de redes têm sido discutidos com relação ao número de participantes, intermitência de suas atividades e longevidade da rede (IBID.). A emergência de duas redes topologicamente diferentes é relatada com relação à frequência de interações: elas podem seguir uma lei de potência ou uma distribuição exponencial (ALBERT; BARABÁSI, 2000). Em redes de listas de e-mail, por exemplo, propriedades livres de escala são relatadas com $\alpha = 1.8$ (como em navegação web e empréstimos de biblioteca (VÁZQUEZ et al., 2006) e traços linguísticos diferentes relacionados a ligações fracas e fortes (MAREK-SPARTZ; CHESLEY; SANDE, 2012).

1.1 REDES COMPLEXAS

Embora não aceito universalmente, é lugar-comum definir uma rede complexa como sendo um “grafo com características topológicas não triviais” (NEWMAN, 2010). Podemos adicionar a essa definição que uma rede complexa é um grafo grande (ainda que não haja consenso sobre o que é *grande* nesse contexto) e que é uma representação em grafo de um sistema encontrado em observações naturais, reais ou empíricas. Outra forma de abordar a definição de “redes complexa” é defini-la como um sistema complexo modelado como rede. Essa segunda definição é útil, mas ainda mais problemática, dado que não há consenso acerca do que é um sistema complexo. Mesmo assim, é vantajoso manter em mente que os autores com frequência definem um sistema complexo como um sistema composto de várias partes em que “o todo é mais que a soma de suas partes”. Os pesquisadores seguidamente consideram que sistemas complexos possuem capacidades de “processar informação”, adaptação e reprodução (HOLLAND, 2014).

Um grafo é uma estrutura que consiste de um conjunto de objetos (chamados vértices) e um conjunto de relações binárias ou duais entre esses objetos (chamadas arestas). Um grafo pode ser sem peso e não direcionado (esse é o caso mais simples), com peso e não direcionado, sem peso e direcionado ou com peso e direcionado. As representações mais usuais de grafos (e redes) são por meio de matriz, lista ou representação vértice-aresta. Na representação por matriz, cada elemento a_{ij} é diferente de zero se i é ligado a j ; elementos diferentes de 0 e 1 podem ocorrer em um grafo com peso; grafos não direcionados resultam em matrizes simétricas. Há duas representações usuais em lista, uma lista cada par de vértices conectados, outra apresenta uma lista para cada vértice em que constam todos os vértices conectados a ele (uma lista de listas). Na representação vértice-aresta, cada vértice é representado como um ponto e cada aresta é representada como uma linha entre os vértices correspondentes. A repre-

sentação em matriz é essencial para raciocínios algébricos e para obtenção de métricas enquanto a representação vértice-aresta é importante para ilustração e para orientar intuitivamente a caracterização dos sistemas.

1.1.1 Uma boa justificativa para o uso da teoria das redes complexas

Uma referência usual de grandeza é o número estimado de átomos do universo ($\approx 10^{80}$). Achemos o número de vértices necessários para que tenhamos tal número de redes possíveis. Consideremos o caso mais simples de redes sem peso e não dirigidas. Cada aresta pode existir ou não e com n vértices há no máximo $\frac{n(n-1)}{2}$ arestas. Portanto:



Ou seja, com apenas 24 vértices há mais redes possíveis do que o número estimado de átomos no universo. Devemos também notar que o número de redes possíveis cresce muito rápido com o número de vértices. Essa é uma boa razão para caracterizarmos esses sistemas através de redes paradigmáticas e métricas genéricas para vértices e redes (e arestas, mas é menos comum).

1.1.2 MÉTRICAS BÁSICAS

As métricas seguintes são bastante básicas e padrão. Além disso, usaremos algumas delas na seção seguinte para caracterizar tipos básicos de redes.

- Grau: número de arestas ligadas ao vértice.
- Grau de entrada: número de arestas terminando no vértice.
- Grau de saída: número de arestas que partem do vértice.
- Força: soma dos pesos das arestas ligadas ao vértice.
- Força de entrada: soma dos pesos de todas as arestas terminando no vértice.
- Força de saída: soma dos pesos de todas as arestas que partem do vértice.

- Centralidade de betweenness: fração de caminhos mais curtos (geodésicas) que incluem o vértice.
- Coeficiente de clusterização: fração de pares de vizinhos que são ligados.

Na discussão seguinte também utilizamos o conceito de distância entre pares de vértices, que é o número de arestas entre eles.

1.1.3 Tipos básicos de redes

Redes complexas são com frequência caracterizadas em termos de modelos paradigmáticos. Existem diversos modelos, mas podemos vislumbrar a teoria por meio dos seguintes (COSTA et al., 2007):

- Modelo de Erdős-Rényi: cada par de vértices é conectado com uma probabilidade fixa p . Esse modelo apresenta um grau característico ($n.p$ onde n é o número de vértices), baixa clusterização e baixa média da distância entre vértices.
- Rede espacial, também chamada rede geográfica ou grafo geométrico: vértices são localizados em espaço métrico e a probabilidade de que dois vértices estejam conectados é maior na medida em que a distância entre os vértices é menor. Essas redes apresentam grau característico, alta clusterização e grande média da distância entre vértices.
- Rede de pequeno mundo: definida como rede em que a distância típica entre vértices cresce com o logaritmo do número de vértices enquanto a clusterização média não é baixa (maior que e.g. no modelo de Erdős-Rényi). Um método para se construir uma rede de pequeno mundo é começar por um reticulado regular em que cada vértice é conectado a k vizinhos mais próximos. Cada aresta é então religada com uma probabilidade p . Com valores intermediários de p tais como $0.01 < p < 0.1$, obtemos uma rede tanto com baixa média da distância entre vértices (tal como na

rede de Erdős-Rényi) como com alta clusterização (como na rede espacial). Esse modelo apresenta grau característico.

- Rede livre de escala: uma rede em que a distribuição do grau segue uma lei de potência ($P(k) \sim Ck^{-\alpha}$ onde C e α são constantes, k é o grau e $p(k)$ é a probabilidade de você encontrar um vértice com grau k se o vértice foi escolhido aleatoriamente).
- Outras redes: entre modelos importantes estão as redes exponenciais, redes com estruturas de comunidades e modelos híbridos.

Redes reais muito comumente apresentam propriedades de pequeno mundo e livre de escala. Esse é o caso da maioria das redes e.g. sociais, genéticas e de alimentos (NEWMAN, 2010). Entretanto, deve-se ter cautela quanto a essa afirmação, dado que redes derivadas de sistemas reais dependem profundamente do que é considerado um vértice e uma aresta, i.e. em como o sistema é modelado como grafo. Outra observação importante é que as redes de Erdős-Rényi, i.e. os grafos do modelo de Erdős-Rényi, são com frequência apontadas como redes com propriedades topológicas triviais. Esse modelo é considerado uma paradigmática “rede complexa”, conceito esse muitas vezes definido como grafos com propriedades topológicas não triviais, o que é uma contradição. Assim, o termo redes complexas não é muito bem definido, como ocorre com a *teoria de complexidade* em geral.

1.2 Polissemia e sinonímia

No contexto de redes complexas, as palavras *rede* e *grafo* são muitas vezes usadas indistintamente, embora a palavra *grafo* possa se referir à estrutura matemática de vértices e arestas e a palavra *rede* possa dizer respeito ao sistema real sendo representado como grafo ou ao grafo obtido por meio da representação de um sistema real. Além disso, a palavra *grafo* pode se referir a um tipo abstrato de estrutura de dados em ciência da computação. A nomenclatura de vértices e arestas varia bastante

entre as áreas de estudo (matemática, física, biologia, sociologia etc.). Um vértice pode ser chamado e.g. um nó, um ponto, um agente, um participante. Uma aresta pode ser chamada e.g. uma ligação, uma relação, um vínculo, uma conexão. Importante para a discussão que segue é que a palavra *hubs* designa, por definição, os agentes mais conectados da rede. Há contextos em que a palavra é usada com outros significados. Por exemplo, no contexto do algoritmo Hyperlink-Induced Topic Search (HITS) para atribuir centralidade aos vértices, mais tradicionalmente para páginas web, os hubs são os vértices com maior grau de saída (maior grau de entrada resulta em autoridades).

2. REDES COMPLEXAS PARA REDES SOCIAIS

As redes sociais são muitas vezes abordadas como redes complexas. **A forma como essas redes são obtidas é variada, e exemplificamos com os seguintes casos:**

- **Rede de amizades:** cada vértice é uma pessoa e cada aresta é uma amizade. Há casos na literatura em que as amizades são consideradas direcionais, i.e. uma pessoa A é amiga de B, mas B não é amiga de A (NEWMAN, 2010). Atualmente, as redes de amizades são comumente obtidas através de plataformas de redes sociais como o Facebook, e nestas as amizades são mútuas, engendrando redes não direcionais. As redes de amizade, em geral, não possuem peso.
- **Redes de interação:** cada vértice é um participante e as arestas indicam interações entre os participantes. Um exemplo relevante de redes de interação é obtido por meio de e-mails (em que uma resposta de B para A é considerada evidência de que B absorveu a mensagem de A, então é feita uma aresta de A para B significando que B absorveu a informação de A). Outro exemplo relevante é a rede obtida por intermédio de retweets (encaminhamentos de mensagens de Twit-

ter). Neste caso, se B retwittou uma mensagem de A, é feita uma aresta de A para B.

Há redes que não são obtidas através de plataformas virtuais. Dois exemplos desses casos, recorrentes na literatura de redes complexas, são as redes do clube de Karatê e a rede de relacionamento de casamentos entre famílias nobres do renascimento de Florença. Fundamentalmente, as redes estão presentes em várias escalas de interação humana (e não humana) e não precisam ser constituídas de indivíduos, o que é demonstrado pela rede das famílias nobres de Florença. Outro tipo de redes sociais, que não são exploradas na literatura de redes complexas até o momento é a derivada de estruturas conceituais, como nas ontologias OWL.

3. ESTABILIDADE TOPOLÓGICA E DIFERENCIAÇÃO TEXTUAL

Em trabalhos recentes (FABBRI et al., 2017), foi encontrada uma extrema estabilidade topológica nas redes de interação humana. Esses resultados dependem de uma base não trivial de física estatística, motivo pelo qual nos atemos a uma discussão qualitativa neste artigo. Alguma estabilidade era esperada, dado que essas redes são descritas como livres de escala pelos estudos, o que implica alguma invariância de suas características. Por sua vez, os participantes dessas redes mudam constantemente seus atributos, até mesmo de quantidade de conexões, o que torna menos óbvio que haja forte estabilidade. Esses resultados podem ser resumidos em três pontos (FABBRI, 2017):

- Medidas de estatísticas circulares ao longo de diversas escalas de tempo (e.g. horas do dia, dias da semana) são bastante regulares ao longo de diferentes estruturas e apresentam padrões bem marcados: há uma diminuição de 1/3 ou até 1/2 da atividade nos finais de semana, há um pico de atividade às 11h da manhã, embora 2/3 das atividades se concentrem no período da tarde, a atividade é bastante homo-

gênea ao longo do mês, nos segundos do minuto e nos minutos da hora.

- As componentes principais no espaço das métricas topológicas possuem pouquíssima variabilidade. Em outras palavras, as medidas se combinam de forma constantes e sempre com a mesma proporção.
- As frações de participantes nos setores de hubs, intermediários e periféricos são bastante estáveis ao longo do tempo, principalmente quando consideradas mais de 200 interações para a obtenção das redes. Caso sejam utilizadas várias métricas ao mesmo tempo, esses setores podem ser discernidos, mesmo em redes pequenas advindas de dezenas de mensagens.

Quando analisados os textos dos três setores (hubs, intermediários e periféricos), obtemos resultados notáveis de diferenças entre eles (FABBRI, 2017). Podemos resumir esses resultados assim:

- Os testes de Kolmogorov-Smirnov de duas amostras resultam inequívocos para a rejeição da hipótese de que os textos dos setores são gerados pelos mesmos processos. Aliás, a diferença entre os textos de setores distintos de uma mesma rede é muitas vezes superior à diferença entre setores iguais de redes diferentes.
- Os periféricos empregam mais substantivos enquanto hubs utilizam mais verbos e advérbios. Além disso, a quantidade de adjetivos não muda substancialmente, o que implica uma maior utilização de adjetivos por substantivo nos setores mais conectados. Isso corrobora a hipótese de que os periféricos trazem diversidade ao sistema, enquanto os hubs qualificam os conceitos e sugerem formas de ação sobre eles.
- Sentenças e mensagens maiores (em caracteres, tokens) são usadas por periféricos. Pontuação é mais abun-

dante no texto de periféricos, embora o tamanho da sentença seja maior.

- Os intermediários com frequência apresentam picos em algumas medições relacionadas ao texto (uso de synsets, tamanho de tokens, etc). Isso é relevante, e até mesmo surpreendente, dado que se pode compreender (provavelmente de forma errônea) que os intermediários são vértices entre os hubs e os periféricos. Estes extremos corroboram o entendimento de que os intermediários constituem um conjunto distinto de participantes.

O mais importante desta discussão é apreender que o texto produzido pelos setores é de fato diferente. Além disso, os intermediários, que não receberam muita ênfase nos pontos anteriores, foram notados como bastante diferenciados em visualizações e análises. Em especial, os intermediários com frequência possuem força de saída muito maior do que de entrada, o que é considerado indício de status (e.g. quando o indivíduo escreve, ele recebe várias respostas). Também pode-se notar que resta aos intermediários estruturar a rede, dado que os periféricos e hubs possuem comportamentos triviais: os periféricos interagem uma ou pouquíssimas vezes; os hubs interagem praticamente em todas as oportunidades.

4. ASPECTOS CRÍTICOS

Nesta seção, são consideradas questões trivialmente relacionadas à teoria crítica. Em especial: a classificação de pessoas e a pesquisa de sistemas sociais sob a ótica da física.

4.1 Tipologia de hubs, intermediários e periféricos

A atribuição de um tipo ao participante salienta questões importantes sobre o cânone científico para tipos humanos e o potencial para preconceito e estigma. O setor de hubs, intermediários ou periféricos, para o qual pertence o participante, pode ser considera-

do como implicando um tipo social para esse participante. Neste caso, o tipo de um participante muda tanto ao longo do tempo em uma mesma rede quanto na medida em que diferentes redes são consideradas. Isso é verdade apesar da estabilidade das redes descrita na seção anterior. Portanto, dado que o tipo do participante muda com o tempo e com relação a diferentes estruturas, o potencial para o preconceito dessa tipologia de hubs, intermediários e periféricos, é atenuado. Em outras palavras, um indivíduo é um hub em um número de redes, periférico em outras redes, e mesmo em uma mesma rede ele muito provavelmente muda de tipo ao longo do tempo.

A importância dessa questão pode ser apreendida pela consideração de tipos estáticos derivados através de critérios quantitativos. Por exemplo, em uma lista de e-mails com um número pequeno de participantes, o número de threads é negativamente correlacionado com o número de participantes. Quando o número de participantes excede um limiar, o número de threads tem uma correlação positiva com o número de participantes. A atribuição de tipos para indivíduos, no último caso, possui maior potencial para o preconceito, pois o tipo de participante obtido é estático e falha no reconhecimento de que indivíduos humanos são entidades mutáveis.

Outras observações sobre essa setorialização da rede em hubs, intermediários e periféricos, e os tipos de participantes que ela acarreta, que são consistentes com a literatura: 1) hubs e intermediários possuem atividade intermitente e atividade estável foi encontrada somente em comunidades menores. 2) A estrutura da rede parece influenciada, principalmente por participantes intermediários, pois eles desempenham papéis menos extremos que hubs e periféricos e podem, portanto, conectar-se a setores e participantes de uma forma mais seletiva e explícita.

4.2 Física antropológica

O estudo de sistemas complexos pode ser realizado como uma empreitada da física, principalmente se redes complexas e estatís-

ticas estão em uso. Quando o sistema complexo é constituído por pessoas, questões intrigantes aparecem de diversos campos, como matemática, ética e sociologia. A “física antropológica” é uma abordagem a esses cenários que permite a pesquisa científica enquanto resolve aspectos éticos e morais através do estudo aberto do eu. Ela propõe uma prática transdisciplinar, cuja relevância emana de assuntos antropológicos e físicos, de sistemas constituídos por humanos e de leis naturais (FABBRI, 2015).

Caso eu tome nota do que tenho feito e do que tem acontecido com meu meio, talvez possa usar estas anotações em meus estudos. De fato, a escrita de diários é uma técnica etnográfica comum, usada com frequência na pesquisa científica. Assim, pode ser razoável que eu use minhas anotações, estejam elas em papéis ou em mídia digital. É também razoável que, se um parceiro quer usar suas anotações, ou que eu as use, que elas sejam usadas. Há algumas iniciativas que recebem esse tipo de dados, novamente evocando questões morais e éticas. Uma abordagem útil foi encontrada em recentes esforços civis, governamentais e acadêmicos, e tem sido chamado física antropológica. As características gerais são:

- Exposição do pesquisador ao meio de interesse, como a uma rede social virtual.
- O uso de anotações dessa exposição, sejam elas logs de atividade, redes de amizade ou interação, conteúdos textuais etc.
- Havendo necessidade, expansão das observações para englobar dados abertos ou doados por parceiros.
- Observação de leis naturais na medida em que elas aparecem em estruturas de rede e na linguagem natural. Todos os recursos são mantidos abertos e publicados, incluindo software, dados e escritos.

Esse arcabouço possibilitou empreitadas utilizando Facebook, Twitter, listas de e-mail e plataformas alternativas enquanto alguma

concordância ética foi mantida entre comunidades e pesquisadores. O aspecto aberto da física antropológica facilitou o suporte tecnológico, sobre o qual são dignos de nota:

- Padrões de dados ligados e web semântica. Ao ser usado para pesquisa, caso não seja considerado invasivo, o dado pode ser publicado como triplas RDF e relacionado a ontologias OWL.
- Exploração de recursos para o indivíduo por meio de critérios de redes complexas e mineração de texto. Isso é compreendido como altamente vantajoso para indivíduos e coletivos e para melhoras de navegação de dados ligados.
- Uso extenso de software livre/aberto. Isso ajuda para que resultados e procedimentos sejam compartilhados através de acesso imediato às ferramentas, versões e algoritmos. Streaming de estruturas sociais, que consiste na exposição em tempo real de nossas redes para a transparência de eventos públicos.

A aplicação da física antropológica é feita na pesquisa diária no mundo todo. O uso explícito do termo tem sido observado em circuitos acadêmicos brasileiros desde fevereiro de 2013, com contribuições de físicos, cientistas da computação, antropólogos, cientistas sociais e filósofos. Ainda assim, essa conceitualização é incipiente e em plena discussão, o que deve acarretar em mudanças significativas. Algumas medidas tomadas para adiantamento dessa frente é: a apresentação de um legado coerente de dados participativos ligados; a implementação de técnicas de exploração de recursos para o indivíduo; a implementação de aparatos de visualização para o compartilhamento de compreensões úteis e para relatoria.

5. APARATO EM SOFTWARE

Atualmente, existem vários de programas disponíveis para a análise de redes sociais: certamente há ao menos algumas dezenas de software bastante relevantes e utilizados

continuamente; há ao menos centenas, e talvez milhares, de software menos utilizados, como atesta uma simples busca com o termo “*social network analysis software*”. Dentre eles, o Gephi é um dos o mais utilizados, pois é dedicado à visualização de grafos, é software livre e é multiplataforma. Para a pesquisa mais elaborada, há bibliotecas para diferentes linguagens de programação. Em especial, podemos citar o NetworkX e o Igraph, ambos acessíveis através da linguagem Python. Outras bibliotecas relevantes são: o Graph-tool, para análise de redes em geral, o Graphviz, para desenho de grafos, e o ccNetViz, para visualização de grafos em navegadores usuais (como Firefox ou Google-Chrome) utilizando WebGL para suportar grafos grandes.


Uma lacuna estava sendo preenchida pelo Netvizz, software pelo qual era possível obter redes de amizades e de interações de indivíduos e grupos do Facebook. No momento, devido às limitações dos termos de uso do Facebook, não é mais possível obter essas estruturas pelo Netvizz. Atualmente, para a obtenção dessas redes do Facebook, tanto de amizades quanto de interações, é necessário utilizar técnicas de *raspagem de dados* (*data scrapping*) por meio de agentes automatizados (muitas vezes, chamados de *bots*).

Para a obtenção dos setores de hubs, intermediários e periféricos, implementamos rotinas publicamente acessíveis em Fabbri (2015). Na mesma biblioteca estão recursos para análise de conteúdos textuais e análises estatísticas mais elaboradas que fogem ao escopo deste artigo. Na biblioteca *Social* (FABBRI, 2016) estão rotinas para a tradução de dados do Facebook, Twitter e IRC para dados ligados (RDF). Na biblioteca *Gmane* (FABBRI, 2016) estão rotinas para tradução de dados de listas de e-mail do Gmane para dados ligados. Em *Participation* (FABBRI, 2016) estão rotinas para tradução de dados das instâncias de participação social ParticipaBR, Cidade Democrática e AA para dados ligados. Para facilitar a renderização audiovisual desses dados, há rotinas para síntese de som e música em *Music* (FABBRI, 2016) e rotinas para visualização

de grafos em evolução através de animações estão sendo migradas para a biblioteca *Visuals* (FABBRI, 2016). Esse aparato em software possibilitou pesquisas por mim e tem o potencial de ser útil para outras pessoas, inclusive pesquisadores. O maior desafio no momento é deixar o legado simples para compreensão e uso, pois há várias implementações entre o que disponibilizamos.

5.1 Relatos do uso desse aparato

Este legado de conhecimento e software possibilitou experimentos de coleta e difusão de informações. Há basicamente dois tipos de experimentos:

- Processos graduais. O experimento mais profundo que realizamos nesta linha consistiu em ativar participante por participante de minha rede de amizades do Facebook, dos mais periféricos (i.e. dos que possuem menos amigos em comum comigo) aos hubs mais conectados (i.e. aos que têm mais amigos em comum comigo). Isso foi feito três vezes, resultando em  5% de amizades a mais a cada ciclo e uma arrecadação de 3000 reais. O mais importante desses ciclos foi o estabelecimento de colaborações com outros pesquisadores e pessoas engajadas na transformação da sociedade, resultando em visitas, artigos e participações em conferências, e uma consultoria da ONU em parceria com a presidência da República (FABBRI, 2014).
- Processos efêmeros. Esses experimentos são caracterizados por serem interações rápidas com as redes. Um exemplo de interesse resultou da observação de que, nas redes de amizade do Facebook que foram analisadas, o conjunto dos 50 participantes de maior centralidade de betweenness era disjunto ao conjunto dos 50 participantes com maior centralidade de closeness (mede o quão perto o vértice está de todos os outros). Esse

fato é bastante inesperado, e pedi a diversos parceiros que mandassem uma mesma mensagem (que fosse do interesse de cada um) para esses dois grupos separadamente. O resultado foi sistemático: o conjunto de maior betweenness reagia sempre amistosamente e compartilhando as mensagens. O conjunto de maior closeness reagia sempre sem responder e com vários membros saindo do grupo de discussão. Nossa hipótese, até o momento, é de que o conjunto de maior betweenness possui algum controle sobre a informação que circula na rede, já o de maior closeness sofre bastante influência da rede mas não tem controle. Esse seria o motivo da disparidade entre as reações dos dois conjuntos de participantes.

Esse repertório de dados e programas também possibilitou a observação da estabilidade dessas estruturas sociais e da diferenciação do texto produzido pelos setores básicos, como já descrito na seção 3. Outro avanço significativo foi o levantamento da nuvem brasileira de dados ligados participativos, descrito em Fabbri (2016). As visualizações até o momento são significativas através do Gephi, do Graphviz e de animações que realizamos para observar a evolução dessas redes de interação, descritas em Fabbri (2013). No entanto, não encontramos facilidade no uso dessas tecnologias por outros pesquisadores. Em parte isso se deve ao fato de que o campo é incipiente, mas também à teoria não trivial que é necessária para o uso das rotinas. Não menos importante é a dificuldade no uso das linguagens de programação, pois a fração das pessoas que escrevem software com fluência ainda é pequena. Há, no entanto, uma abundância de recursos básicos para análise de redes complexas em pleno uso e desenvolvimento por comunidades bastante consideráveis, como o NetworkX e o Gephi. Há maiores dificuldades para reutilização de rotinas mais elaboradas que não se atêm à obtenção de medidas ou visualização. Uma exceção a esse

respeito é que as ferramentas em uso possuem capacidades para detecção de comunidades (NEWMAN, 2006).

6. CONSIDERAÇÕES FINAIS

Este ensaio traz de forma bastante simples e instrumental a área de redes complexas por meio de discussões conceituais, medidas, paradigmas e resultados recentes sobre redes sociais. Há também a relação entre topologia e texto, o que é incomum na área, e exibimos conceitos potencialmente novos como “física antropológica” e a obtenção de tipos humanos hubs, intermediários e periféricos por critérios quantitativos. Igualmente relevante para a abordagem do assunto, é apresentado aparato em software de terceiros e desenvolvido com a nossa pesquisa, para análise topológica, de texto, obtenção de dados ligados e visualização. Alguns resultados do uso desse arcabouço na seção 5 devem facilitar o leitor a compreender a área e o instrumental de forma mais profunda e para que possa trazer novas contribuições.

Nossos trabalhos futuros devem envolver um aprofundamento de conceitos como

física antropológica e dos tipos humanos derivados dos setores da rede. A melhoria dos software que disponibilizamos é contínua e deve acarretar em melhorias nas bibliotecas atuais e em novas bibliotecas. Uma demanda já identificada é a disponibilização de rotinas para raspar os dados de redes sociais como o Facebook, que não permite um acesso fácil a essas estruturas. Já os programas que não escrevemos podem contar com mais empenho de nossos pesquisadores para seus desenvolvimentos, em especial o Graph-tool parece ter bastante empenho de brasileiros. Já os resultados de estabilidade topológica e diferenciação textual necessitam de maiores empenhos dos pesquisadores já envolvidos pois, dada a complexidade do assunto e especificidade dos métodos, é improvável que novos pesquisadores se aventurem profundamente nessas questões em um futuro próximo. Uma linha de estudos que temos discutido há alguns anos é dedicada à obtenção de tipologias de indivíduos e de redes através de dados de redes virtuais e de métodos quantitativos. Tais tipologias serão influenciadas pela teoria de tipos autoritários (ADORNO et al., 1950).

REFERÊNCIAS

- ALBERT, R.; BARABÁSI, A.-L. Topology of evolving networks: local events and universality. **Physical Review Letters**, v. 85, n. 24, p. 5.234, 2000.
- BALL, B.; NEWMAN, M. E. **Friendship networks and social status**, 2012. Disponível em: <https://arxiv.org/pdf/1205.6822.pdf>. Acessado em: 23 jan. 2016.
- BIRD, C. et al. **Mining email social networks**. In: INTERNATIONAL WORKSHOP ON MINING SOFTWARE REPOSITORIES, 2006, Shanghai. Proceedings... New York: ACM, p. 137-143, 2006.
- FABBRI, R. **Python package to observe temporal stability in the GMANE database**, 2015. Disponível em: <https://github.com/ttm/percolation>. Acessado em: 10 mar. 2016.
- FABBRI, R. **Python package to render linked data from Facebook, Twitter and IRC**, 2015. Disponível em: <https://github.com/ttm/social>. Acessado em: 10 mar. 2016.
- FABBRI, R. **Python package to render linked data from ParticipaBR, Cidade Democrática and AA**, 2015. Disponível em: <https://github.com/ttm/participation>. Acessado em: 10 mar. 2016.
- FABBRI, R. **Python package to render linked data from Gmane e-mail lists**, 2015. Disponível em: <https://github.com/ttm/gmane>. Acessado em: 10 mar. 2016.
- FABBRI, R. **Content extraction through API from the Brazilian federal portal of social participation and its tools to a social participation cloud**, 2014. Disponível em: <https://github.com/ttm/pnud5/blob/master/latex/produto.pdf?raw=true>. Acessado em: 30 out. 2016.

FABBRI, R. **What are you and I?** [Anthropological physics fundamentals]. 2015. Disponível em: https://www.academia.edu/10356773/What_are_you_and_I_anthropological_physics_fundamentals_. Acessado em: 23 jan. 2016.

FABBRI, R. **Ensaio sobre o auto aproveitamento: um relato de investidas naturais na participação social**, 2014. Disponível em: <https://arxiv.org/pdf/1412.6868.pdf>. Acessado em: 10 mar. 2017.

FABBRI, R. **A connective differentiation of textual production in interaction networks**, 2013. Disponível em: <https://arxiv.org/pdf/1412.7309.pdf>. Acessado em: 10 mar. 2017.

FABBRI, R. et al. **Temporal stability in human interaction networks**, 2017. Physica A. Available from: <https://arxiv.org/pdf/1310.7769.pdf>. Accessible at: 23 jan. 2016.

FABBRI, R.; OLIVEIRA JUNIOR, O. N. **Linked open social data for scientific benchmarking**. 2016. Disponível em: <https://github.com/ttm/linkedOpenSocialData/raw/master/paper.pdf>.

LATOUR, B. Reassembling the social: an introduction to actor-network-theory. **Journal of Economic Sociology**, v. 14, n. 2, p. 73-87, 2013.

LEICHT, E. A. et al. Large-scale structure of time evolving citation networks. **European Physical Journal B**, v. 59, n. 1, p. 75-83, 2007.

MAREK-SPARTZ, K.; CHESLEY, P.; SANDE, H. **Construction of the gmane corpus for examining the diffusion of lexical innovations**, 2012. Disponível em https://pdfs.semanticscholar.org/03c0/8a95695e39595267f786a1040cd002bfea1d.pdf?_ga=1.212330838.358458694.1488839238. Acessado em: 23 jan. 2016.

MORENO, J. L. **Who shall survive?: a new approach to the problem of human interrelations**. Washington: Nervous and Mental Disease Publishing Co, 1934 (Nervous and mental disease monograph series, n. 58).

NEWMAN, M. **Networks: an introduction**. Oxford: Oxford University Press, 2010.

NEWMAN, M. E. Modularity and community structure in networks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 23, p. 8.577-8.582, 2006. ONNELA, J.-P. et al. Geographic constraints on social network groups. **PloS One**, v. 6, n. 4, p. e 16939, 2011.

PALCHYKOV, V. et al. Sex differences in intimate relationships. **Scientific Reports**, v. 2, 2012. doi:10.1038/srep00370.

PALLA, G.; BARABÁSI, A.-L.; VICSEK, T. Quantifying social group evolution. **Nature**, v. 446, n. 7.136, p. 664-667, 2007.

TRAVENÇOLO, B.; COSTA, L. F. Accessibility in complex networks. **Physics Letters A**, v. 373, n. 1, p. 89-95, 2008.

VÁZQUEZ, A. et al. Modeling bursts and heavy tails in human dynamics. **Physical Review E**, v. 73, n. 3, p. 036127, 2006.

DADOS DO AUTOR

RENATO FABBRI

Doutor em física computacional no Instituto de Física de São Carlos, Universidade de São Paulo. Pesquisa e desenvolve software com ênfase em física, computação, matemática, artes, tecnologias e conhecimentos livres. São Carlos/SP – Brasil. renato.fabbri@gmail.com

Submetido em: 7-4-2017

Aceito em: 29-8-2017