

# Analítica Visual (Visual Analytics) de Topologia e Texto em Redes Sociais

**Proponente:** Dr. Renato Fabbri

**Supervisora:** Profa. Dra. Maria Cristina Ferreira de Oliveira  
Instituto de Ciências Matemáticas e de Computação,  
Universidade de São Paulo (ICMC-USP)

29 de maio de 2017

**Resumo:** Este projeto propõe o desenvolvimento de métodos e software de analítica visual (*visual analytics*) para a análise de redes sociais (ARS). As contribuições visam permitir a utilização de interfaces visuais interativas por pesquisadores de diversas áreas para a classificação/tipologização de redes e participantes em dados relacionados a redes sociais virtuais. Uma caracterização destes sistemas, por critérios topológicos e textuais, assim como visualizações preliminares, e o levantamento de uma base de dados em protocolo conveniente (dados ligados), foi realizada pelo proponente em seu trabalho de doutorado. Além disso, o proponente possui formação multidisciplinar e experiência em programação web e multimídia, o que deverá favorecer o bom andamento da pesquisa. A supervisora deste projeto possui ampla experiência em visualização de dados e tem investido em analítica visual e análise de redes sociais. Todo o trabalho aqui proposto deverá ser conduzido de forma aberta, todo o código desenvolvido publicamente em repositórios Git, e acompanhado de escrita de artigos e parcerias com pesquisadores de outras áreas.

**Palavras-chave:** Visual analytics, Redes sociais, Redes complexas, Mineração de texto.

# Visual Analytics of text and topology in social networks

**Proponent:** Dr. Renato Fabbri

**Supervisor:** Prof. Dr. Maria Cristina Ferreira de Oliveira

Institute of Mathematics and Computational Sciences,

University of São Paulo (ICMC-USP)

29 de maio de 2017

**Abstract:** This project proposes the development of visual analytics methods and software for social networks analysis (SNA). The contributions aim to allow the use of interactive visual interfaces by researchers from different areas for the classification/typologization of networks and participants in data related to virtual social networks. A characterization of these systems, by topological and textual criteria, as well as preliminary visualizations, and the collection of a database in linked data, was carried out by the proponent in his doctoral work. In addition, the proponent has multidisciplinary training and experience in web and audiovisual programming, which should favor the good progress of the research. The supervisor of this project has extensive experience in data visualization and has invested in visual analytics and social network analytics. All work proposed here should be openly conducted, all developed code publicized in Git repositories, and accompanied by scientific writing and partnerships with researchers from other areas.

**Keywords:** Visual analytics, Social networks, Complex networks, Text mining.

# 1 Enunciado do problema

Este trabalho se propõe a desenvolver métodos e ferramentas para a Análise de Redes Sociais (ARS) via Analítica Visual (*visual analytics*), orientados a especialistas de outras áreas (e.g. linguistas, físicos, cientistas sociais, psicólogos) e considerando as propriedades das estruturas sociais em análise. Estas propriedades podem ser divididos em três grupos:

- Topológicos: as redes sociais são reportadas como livres de escala, de pequeno mundo e com comunidades [1].
- Textuais: a linguagem nas redes sociais é muitas vezes informal, com abreviações e erros com relação às normas oficiais [2].
- Relacionamento entre topologia e texto: há aspectos do texto produzido pelos participantes que se relacionam à topologia. Por exemplo, *hubs* utilizam frases mais curtas e mais verbos e advérbios, participantes periféricos utilizam mais substantivos [3].

Em resumo, nossa proposta de desenvolver métodos e uma ferramenta de analítica visual para redes sociais envolve a integração dos dados a serem analisados, o processamento dos dados para obtenção de estruturas relevantes, a concepção de visualizações e mecanismos de interação com elas, e a escrita de software que permita aos pesquisadores de outras áreas usufruir destas rotinas por meio de uma interface unificada.

## 1.1 Conceitualização, terminologia, polissemia e sinônimos

A pesquisa proposta neste documento envolve várias áreas do conhecimento. O conhecimento pode ser sistematizado de diversas formas. Entendemos que a divisão a seguir é útil para o trabalho:

- Analítica Visual (*visual analytics*): compreendida aqui como a mineração de dados apoiada por visualização de dados em interfaces interativas. Uma definição mais usual é: “a combinação de técnicas automáticas de análise com visualizações interativas para a compreensão e tomada de decisão em bases de dados grandes e complexas” [4]. A Figura 1 explicita a relação entre mineração de dados e a analítica visual. Embora a área seja relativamente recente, já existem algumas diretrizes consagradas pela literatura para abordar a analítica visual. Por exemplo, há o lema/mantra de Shneiderman “Visão geral primeiro, zoom e filtragem, detalhes sob demanda” [5], e uma versão estendida dele “Analise primeiro, mostre o importante, zoom e filtragem, analise mais, apresente detalhes sob demanda” [6].

- Análise de Redes Sociais (ARS): concebida nesta proposta a partir das áreas de redes complexas e mineração de texto.

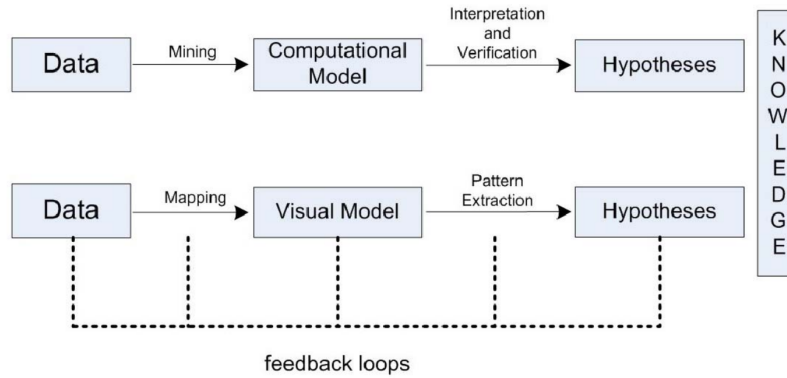


Figura 1: Típico diagrama expondo as relações entre a analítica visual (abaixo) e a mineração de dados (acima). As questões principais são: 1) na analítica visual há a inserção do usuário em várias etapas da aquisição de conhecimento a partir dos dados; 2) na analítica visual há a utilização da visualização de dados pelo usuário para a mineração. Como pode-se inferir, a analítica visual implica na utilização de interfaces visuais interativas para representação dos dados que possam ser adaptadas continuamente no processo de obtenção e evidenciação de hipóteses.

Fonte: Eduard Gröller [7].

Salta evidente que as duas áreas são multidisciplinares. As demais áreas diretamente relacionadas pela nossa descrição são também multidisciplinares e também recentes: mineração de dados, visualização de dados, interfaces interativas, redes complexas, mineração de texto. As redes complexas surgiram como área de investigação reconhecida no final da década de 90 [1]; os primeiros workshops de mineração de texto ocorreram em 1999 [8]; como área independente, o desenho de grafos (*graph drawing*) surgiu na década de 1990 [9]. Ainda não citado, mas importante para a forma como abordamos alguns dos desafios da pesquisa proposta, o termo “dados ligados” foi cunhado em 2006 [10]. Como resultado, o vocabulário não é sempre bem definido, há termos diversas que se referem aos mesmos conceitos e conceitos diferentes relacionados a um mesmo termo.

Uma apresentação detalhada da terminologia, incluindo escolhas consistentes de termos e os diferentes problemas nas terminologias encontradas, foge ao escopo deste projeto de pesquisa. Para exemplificar, na área de redes complexas, os termos “rede” e “grafo” são utilizados muitas vezes de forma indistinta, embora em textos mais cuidadosos a palavra grafo se refira à estrutura matemática composta de vértices e arestas (relacionamentos binários/duais entre vértices) e

rede em geral se refira a grafos encontrados em sistemas reais/empíricos/naturais ou ao sistema representado como grafo. Outro bom exemplo é a usual indistinção entre mineração de texto, linguística computacional e processamento de linguagem natural (PLN). Uma distinção coerente entre as áreas, não encontrada na literatura por nós, pode ser estabelecida considerando a mineração de texto como a mineração de dados aplicada a textos, a linguística computacional como a modelagem (estatística ou orientada a regras) de linguagem natural, e o PLN como a pesquisa e desenvolvimento aplicados à interação entre computadores e a linguagem natural.

Outros termos e áreas importantes para o trabalho proposto são: interação humano-computador, percepção, cognição, estímulo, representação, representação de dados (de uma, duas, três ou mais dimensões), visualização (de dados, de informação, científica), raciocínio analítico (*analytical reasoning*), reconhecimento de padrões, aprendizado de máquina, design, design de interação, design de interfaces, cores, formas, glifos, analítica visual (*visual analytics*), análise de tipos específicos de dados (e.g. áudio, imagem, estruturas sociais, dados sintéticos).

Talvez a conceitualização mais útil para esta pesquisa, que transcende o propósito de expor inconsistências ou a busca vocabulários e definições concisos, é o estudo da classificação. Nas ciências humanas, há tradição no estudo de *tipologias*, que são classificações em geral não bem definidas quantitativamente e mais preocupadas em qualificar os diferentes tipos ou classes. Já as *taxonomias* são em geral reconhecidas como de grande tradição na biologia, e possuem ênfase na nomenclatura com base em características bem definidas. As definições de classificação, taxonomia e tipologia não são uniformes na literatura científica (e.g. árvore taxonômica na linguística se refere ao relacionamento dos hiperônimos aos hipônimos e não se referem à biologia) e há outros termos referentes à classificação: vocabulários, ontologias, tesouros, etc. Ainda assim, a consideração da literatura sobre classificação em diferentes áreas do conhecimento poderá ajudar-nos a atingir uma melhor organização do conhecimento e a contemplar as necessidades de pesquisadores com formações diversas.

## 1.2 Sua importância

Há uma proliferação de redes sociais virtuais que produzem muitos dados e estas estruturas estão ainda sendo caracterizadas. Uma ferramenta de analítica visual em código aberto, utilizável em sistemas operacionais diferentes, e que seja capaz de processar grandes quantidades de dados é certamente algo desejável. As análises a serem possibilitadas pela ferramenta, e necessárias para a concepção da mesma, são úteis para aprofundar a nossa compreensão sobre as redes sociais,

principalmente as virtuais.

Na pesquisa de doutoramento do candidato, foram feitas contribuições para a caracterização destas estruturas sociais a partir das áreas de redes complexas e de mineração de texto. O trabalho aqui proposto visa aplicar os mesmos métodos a um conjunto maior de redes virtuais para melhor delimitar a validade da caracterização e para aprofundar as análises. Este conhecimento sobre as estruturas analisadas possui um potencial para fundamentar classificações e tipologias nas redes e nos participantes. Além da relevância evidente das redes sociais virtuais, há interesse neste estudo de tipos (de redes e de participantes) de nossa parte, da parte da Rede Nexos de Pesquisa Interdisciplinar [11] (do qual o proponente é membro há anos) e da tradição de exatas em geral, que coloca grande ênfase na tarefa de classificação.

Melhor relacionando ARS com redes complexas e mineração de texto, podemos esperar que haja contribuições na análise de literatura e de redes livres de escala. Este aspecto não é central para a proposta aqui apresentada, mas ainda assim é relevante pois, para melhor permitir o estudo das redes sociais, é útil que haja outras redes e textos para comparação, o que permitirá a ARS em relação a estas estruturas externas e vice-versa.

### 1.3 A contribuição se bem sucedido

Do ponto de vista dos métodos, as análises pertinentes, assim como escolhas de interfaces apropriadas, já resultaram em publicação científica pelo proponente [12] e pela supervisora [13, 14] e deverá implicar em maiores contribuições para a compreensão da analítica visual e das estruturas sociais virtuais. O software a ser desenvolvido certamente é uma contribuição pois facilitará análises por pesquisadores de outras áreas, avanços na analítica visual e na caracterização dos sistemas sociais. Potencialmente também será facilitada a análise de literatura e de redes livres de escala em geral.

### 1.4 Trabalhos relevantes

A tese de doutoramento do proponente *“Topological stability and textual differentiation in human interaction networks: statistical analysis, visualization and linked data”* apresenta contribuições para a caracterização dos sistemas sociais virtuais, algumas destas em processo de publicação [15, 16, 17] e uma já aceita pela revista *Physica A* [12]. O proponente deste projeto de pós-doutoramento concebeu e implementou uma técnica de visualização de redes em evolução [3, 18], ilustrado na Figura 2. Há trabalhos do grupo de pesquisa *Visualization, Imaging and Computer Graphics* (VICG/ICMC/USP) em visualização de dados textuais

e de redes [19, 20, 21, 22, 23]. Os textos [24, 25] são base para visualização de dados, enquanto [26, 27, 28] são especificamente para visualização de redes/grafos e [29, 30] são sobre mineração de texto. Os trabalhos [31, 4] são bastante relevantes para a analítica visual. Há diversos software que deverão ser utilizados no decorrer do trabalho para apreender as possibilidades de analítica visual neste contexto (e.g. Gephi, Weka, ccNetViz, D3.js, Three.js). Há também literatura e artefatos mais ligados à arte, como a teoria da Gestalt, kiki-bouba, design e estética [32, 33, 34, 35].

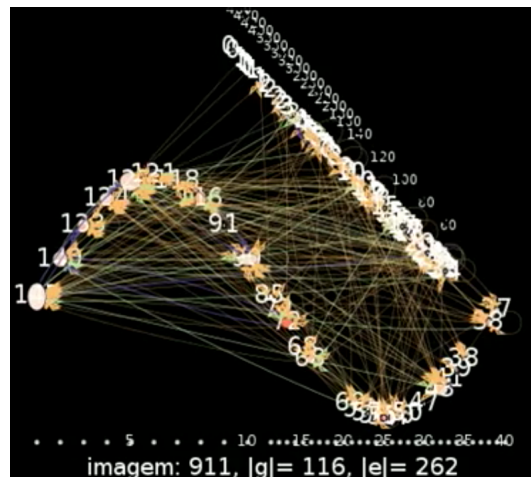


Figura 2: Um quadro do método de visualização de redes em evolução via animações (chamado Versinus [18]). Os vértices mais conectados (hubs) são posicionados na metade do período da senoide à esquerda. Os vértices intermediários são posicionados na segunda metade da senoide, enquanto os periféricos estão na linha reta superior. Os pontos brancos na parte inferior, com numeração, mantêm a posição do vértice na ordenação global. Valores de medições piscam periodicamente perto dos vértices aos quais elas estão relacionadas.

Fonte: figura feita pelo proponente.

#### 1.4.1 Projeto de Pesquisa Regular relacionado a esta proposta de pós-doutoramento

A Profa. Dra. Maria Cristina Ferreira de Oliveira, supervisora desta proposta, está pesquisando analítica visual para redes sociais e possui um aluno de mestrado no tema. Ela enviou recentemente uma proposta de projeto de pesquisa à FAPESP (Processo 2017/05838-3). Segue o resumo.

“A pesquisa em Visual Analytics é central no tratamento dos desafios associados à análise de dados e computação intensiva em dados, pelo potencial de combinar técnicas de Aprendizado de Máquina e de Visualização para apoiar a interpretação de dados complexos. O acoplamento de técnicas oriundas de ambas

as áreas pode promover avanços significativos na capacidade humana de análise de dados, pois permite a indivíduo e computador assumirem papéis complementares ao tratar os muitos problemas introduzidos pelo volume e complexidade dos conjuntos de dados gerados em diversos domínios de aplicação. Este projeto de pesquisa aborda dois focos distintos em visual analytics, um de natureza aplicada e outro de natureza conceitual. No aspecto aplicado serão considerados (i) o problema de visualização de redes de grande escala, com ênfase em redes sociais; e (ii) o problema da análise exploratória de espaços de atributos que caracterizam fenômenos multivariados e variantes no tempo - por exemplo, resultantes de sensores utilizados para monitoramento ambiental em diversos domínios. Em ambos os casos, a busca por soluções escaláveis para grandes volumes de dados representa um desafio. No aspecto conceitual, dando continuidade a uma colaboração em andamento, iremos conceber e realizar alguns estudos experimentais que contribuam para esclarecer os processos cognitivos subjacentes à interpretação de um tipo particular de visualização multidimensional, os chamados mapas de similaridade. A análise dos resultados pode sugerir modelos conceituais sobre a interpretação desse tipo de mapeamento visual. Esperamos com esse estudo contribuir para ampliar o embasamento conceitual sobre essas técnicas, essencial para futuros avanços na área.”

Ou seja, há um interesse do grupo VICG (*Visualization, Imaging and Computer Graphics*) na ARS através da analítica visual que é anterior à concepção desta proposta de pós-doutoramento. De forma bastante conveniente, o trabalho de doutoramento do candidato fornece um arcabouço de análise, dados já representados como dados ligados e visualizações iniciais, o que favorece a sinergia entre esta proposta de pós-doutoramento, e a pesquisa em andamento no VICG.

#### **1.4.2 Painel de Gestão Acadêmica a partir do ScriptLattes**

Dada a sobreposição de técnicas entre o projeto aqui proposto e as utilizadas no “Painel de Gestão Acadêmica”, promovido pelo Centro de Competência em Software Livre (CCSL-ICMC/USP), a partir do ScriptLattes [36], consideramos colaborar com esta frente de pesquisa e desenvolvimento. Esta possibilidade é reforçada pelo fato de que se dedicam ao painel um número de pesquisadores do ICMC/USP, instituição na qual a supervisora desta proposta de pós-doutoramento é professora titular.

## **2 Resultados esperados**

São esperados os resultados:



- melhor caracterização das redes sociais virtuais através de mineração da topologia e do texto.
- Avanços em analítica visual para redes sociais principalmente através da consideração das características topológicas e textuais para a interface e pré-processamento dos dados.
- Um software aberto para analítica visual de redes sociais que seja capaz de lidar com grandes volumes de dados e seja executado em navegadores web.
- Um arcabouço de analítica visual que permita comparar redes sociais com outras redes reais (especialmente livres de escala) e com textos diversos (e.g. literatura científica, James Joyce, Shakespeare, Bíblia).
- Um arcabouço de analítica visual que facilite estudos classificatórios (tipológicos, taxonômicos) em redes sociais de forma a permitir a observação de tipos/classes de redes e de participantes por pesquisadores e.g. de ciências e psicologia sociais.

## 2.1 Disseminação dos resultados

O software aberto pretendido deve facilitar a interação com usuários, desenvolvedores e pesquisadores. Visamos que o software seja utilizado por pesquisadores de áreas diversas como linguística, física e ciências sociais. Os resultados devem ser apresentados em conferências e em artigos de revistas, principalmente internacionais. A circulação de material didático em vídeo potencialmente aumentará a recepção das contribuições. As análises e interfaces devem ser devolvidas às comunidades virtuais que geraram os dados e os software utilizados.

## 3 Desafios científicos e tecnológicos

A consideração da tarefa revela alguns desafios. Primeiro, a análise de texto e de topologia implica em um conjunto de medidas de alta dimensionalidade. Segundo, o montante de dados a serem analisados é grande e as medições são muitas vezes computacionalmente custosas. Por exemplo, a medida usual de *betweenness centrality* requer que sejam encontradas todas as geodésicas entre todos os pares de vértices [37]; as medidas de texto são resultantes de muitos dados e podem envolver procedimentos não triviais como a etiquetagem morfosintática ou similaridades via rede semântica (Wordnet) [38].

Este projeto propõe a utilização destas análises, e aprofundamento delas, para a analítica visual. Os desafios imediatos para esta tarefa são:

- O uso apropriado de técnicas de visualização. Os layouts para redes orientados a força são os mais usuais em análise de redes sociais por explicitarem comunidades e resultarem em figuras esteticamente atraentes, mas a observação de outras características ficam prejudicadas, um fenômeno que tem recebido o nome de *hairball effect*. Para a visualização de dados multidimensionais, as técnicas podem não ser apropriadas para dados com dimensionalidade muito alta ou as não resultarem em coordenadas com fácil interpretação, como no caso de algumas projeções multidimensionais (e.g. LSP). Além disso, a escolha de cores, formas e glifos possui um alto impacto na capacidade de análise pelo usuário e possui muitas teorizações envolvidas e são substancialmente modificadas de acordo e.g. com o monitor do usuário, o que dificulta o controle na implementação. Para exemplificar este aspecto desafiador do trabalho proposto, na Figura 3 está uma sistematização bastante resumida de parte das questões envolvidas apenas na visualização de grafos.
- Viabilizar interatividade apropriada com o montante de dados e medições exposto acima. Por exemplo, quais controles devem ser disponibilizados e com quais controladores? Como compatibilizar a interatividade e a grande quantidade de dados a serem analisados?
- Possibilitar a análise de estruturas arbitrárias, e.g. de redes formadas por quantidades arbitrárias de mensagens de emails ou da junção de diferentes redes do Facebook.
- A escrita de software multiplataforma que lide com a integração dos dados, processamento destes para obtenção de estruturas relevantes, visualização e interatividade.

### 3.1 Meios e métodos para superá-los

Desafios da análise de topologia e texto em redes sociais foram abordados pelo proponente em [3], resultando em uma caracterização das estruturas sociais quanto à estabilidade topológica e diferenciação textual. Para os outros desafios ligados explicitamente à analítica visual, itemizados acima, propomos:

- A utilização de layouts geometricamente inspirados, como os *Hive Plots* [39] ou o Versinus (este desenvolvido pelo proponente [3]). A possibilidade de

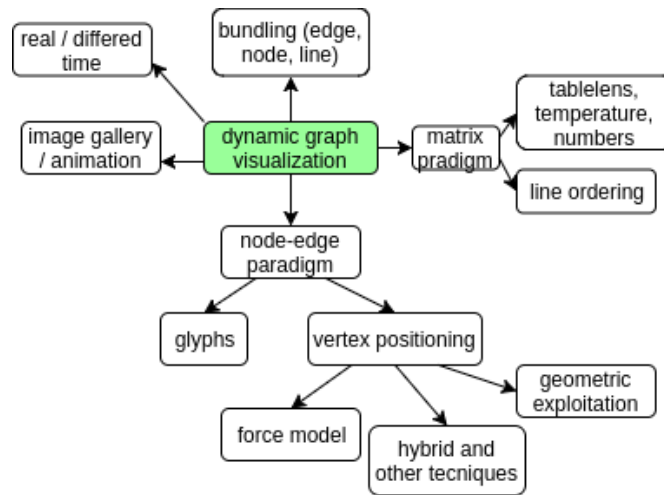


Figura 3: Resumo de opções relacionadas à visualização de grafos. Uma visualização de grafo em evolução pode ser feita por imagens ou animação, pode ser feita em tempo real ou diferido, pode envolver empacotamento de vértices ou arestas, pode ser baseada na representação do grafo por matrizes ou via diagramas nó-aresta. Neste último caso, há no mínimo que se considerar os glifos usados para os vértices e os posicionamentos destes. As setas neste diagrama significam “pode ser” ou “pode usar”.

Fonte: figura feita pelo proponente.

utilização de layouts e plots diferentes e de layouts simulâneos em regiões distintas da interface. Há também a previsão do uso de técnicas de redução de dimensionalidade (e.g. MDS, LSP) dada a grande quantidade de medidas relevantes para analisar redes sociais quanto à topologia e ao texto. Outras visualizações já bem estabelecidas devem ser disponibilizadas e.g. plot quantil-quantil (Q-Q) e de cumulativas de distribuições para a análise de dados de natureza estatística.

- O estudo da literatura da área nos fornece pistas valiosas, mas certamente só saberemos quais as reais possibilidades de interatividade através da implementação e de testes subsequentes. Prevemos também que, evidentemente, as possibilidades de interatividade aliadas ao processamento massivo de dados dependerá do hardware do usuário, uma vez que deverá ser executado no cliente ao menos a visualização.
- Para que as estruturas a serem analisadas possam ser integradas e observadas por partes, devemos aproveitar a representação de estruturas sociais como dados ligados já desenvolvida pelo proponente [3, 17]. A base de dados resultante (já obtida) preferencialmente deve ser abrigada em alguma instância publicamente acessível e pode ser expandida no decorrer da pes-

quisa. Uma possibilidade gratuita é utilizar os serviços do Data-World [40] pois abrigam dados ligados, mesmo na grande quantidade que temos, e permitem consultas via SparQL.

- Para portabilidade e fácil utilização, entendemos que o ideal é que o software possa ser executado em navegadores web usuais (e.g. Google Chrome e Firefox). Para isso, contamos com boas bibliotecas de visualização. A principal é a D3.js, que pode ser integrada com WebGL para possibilitar visualizações mais complexas. Boas alternativas são o Three.js e o ccNet-Viz, que já são integradas com WebGL. Já no processamento dos dados, podemos aproveitar o bom repertório de bibliotecas em Python, que utiliza as rotinas em Fortran do LAPACK para processamento matricial e vetorial. Devemos estudar as possibilidades de implementar o processamento dos dados diretamente em JavaScript para que a arquitetura do software fique simplificada, mas isso não é central, dado que a integração de JavaScript com Python é imediata caso sejam utilizados frameworks em Python para desenvolvimento web como Flask e Django. A maior diferença neste caso é que a implementação das rotinas de processamento dos dados em Python serão executadas pelo servidor ao passo que, se implementadas em JavaScript, poderão ser executadas no cliente (caso seja feita uma implementação em Node.js, o processamento em JavaScript poderá ser feito no servidor). Concebemos toda a implementação em código aberto/software livre para facilitar a colaboração no desenvolvimento e permitir que as rotinas sejam publicamente escrutináveis.

Para o desafio da organização do conhecimento, prevemos que será importante que a interface seja capaz de registrar e recuperar configurações que os usuários acharem relevantes. Aliado a isso, a interface deve ser capaz de guardar anotações relacionadas às visualizações, como sobre levantamento de hipóteses ou evidências de confirmações das mesmas.

## 3.2 Adequação do candidato à proposta

O candidato a esta pesquisa possui formação multidisciplinar, o que deverá contribuir para enriquecer o escopo do projeto. Expandindo o argumento:

- a pesquisa envolve muitas áreas de conhecimento, de questões ligadas ao audiovisual e percepção, a análise de texto e redes e engenharia de software, como exposto na Seção 1.1.
- O candidato desenvolveu trabalhos nas áreas envolvidas: é graduado em artes (composição musical); realizou mestrado e doutorado na física com-

putacional com contribuições científicas de nítido cruzamento entre física, estatística, mineração de texto, redes complexas, ontologias e dados ligados, programação, percepção e artes [41, 3, 12, 16, 18, 17, 42, 43, 44]. Além disso, realizou diversas atividades de cunho prático, como composição musical, apresentações artísticas, programação de software original e contribuição para diversos software livres bastante utilizados no mundo todo, e uma consultoria em parceria com a ONU e a Presidência da República [45, 3, 41].

Em resumo, o candidato está habituado a lidar com as diferentes linguagens e a grande quantidade de literatura envolvida em um trabalho como este. Adicionalmente, possui algum conhecimento e maturidade com relação às áreas, como evidenciado pelo método de visualização de redes em evolução que desenvolveu [18], pela mineração de dados realizada no seu doutorado [3], a modelagem psicofísica apresentada em seu mestrado [41, 46], e as implementações computacionais que permeiam praticamente toda a sua produção.

## 4 Cronograma

Este projeto foi dividido nas seguintes etapas, cujo cronograma de execução é apresentado na Tabela 1.

1. Revisão da literatura referente à analítica visual.
2. Acréscimos aos modelos atuais de analítica visual e visualização de dados aplicados à análise de redes sociais, com o foco no participante da rede, nos pesquisadores em potencial e na classificação/tipologia de redes e participantes.
3. Implementação computacional. Estamos já implementando layouts para grafos no ccNetViz.
4. Escrita e publicação dos resultados em artigos. Esta etapa está já em andamento pois possuímos diversos escritos com resultados relacionados à mineração e visualização de dados de redes sociais que estão sendo submetidos para publicação.
5. Trocas com pesquisadores externos, estabelecimento de colaborações.
6. Elaboração do relatório científico final.

Atividade	2017	2018		2019
	2°	1°	2°	1°
1	•	•		
2	•	•		
3	•	•	•	•
4	•	•	•	•
5	•	•	•	•
6				•

Tabela 1: Cronograma de atividades ao longo dos semestres.

## 5 Disseminação e avaliação

Como exposto na Seção 2.1, haverá disseminação das contribuições em conferências, revistas, vídeo e através da disponibilização do software em código aberto e de parcerias para utilização dele na aquisição de conhecimento e geração de literatura científica. O desenvolvimento do software estará publicamente acessível desde o começo, muito provavelmente em repositório Git na plataforma Github. A avaliação deverá ser feita em três frentes: pela nossa própria capacidade de mineração dos dados, idealmente para obtenção de resultados suficientemente relevantes para publicação; através de testes com usuários em potencial (e.g. da Rede Nexos de Pesquisa Multidisciplinar); através da submissão das contribuições de analítica visual para revistas e conferências, de forma a obter avaliações de revisores especialistas.

## 6 Outros apoios

Para a realização deste projeto, contamos com apoio institucional e infraestrutura do Instituto de Ciências Matemáticas e de Computação (ICMC/USP) e do grupo *Visualization, Imaging and Computer Graphics* (VICG/ICMC/USP). Há um Projeto de Pesquisa Regular em análise pela FAPESP (título: “Visual analytics: aplicações e uma investigação conceitual”) em que consta a analítica visual para redes sociais, o que explicita o interesse da supervisora e do grupo de pesquisa nesta proposta de pós-doutoramento. O proponente deste projeto foi aprovado para uma parceria com a Universidade de Nebraska-Lincoln e o Google no âmbito do programa *Google Summer of Code* (GSoC 2017). A duração desta parceria é curta, de apenas 3 meses, e já está em andamento, e foi compreendida pelo candidato como uma oportunidade de iniciar o trabalho em analítica visual: sua responsabilidade nesta parceria é implementar layouts para visualização de redes no ccNetViz, uma biblioteca para visualização de redes em

navegadores HTML utilizando WebGL. Para utilização dos desenvolvimentos propostos neste projeto, contamos com os demais membros da rede Nexos de pesquisa [11] (grupo de pesquisa cadastrado no CNPq) e com membros do grupo labMacambira.sourceforge.net, dentre os quais estão vários acadêmicos e outros interessados em pesquisas sobre audiovisual, estruturas sociais e desenvolvimento de software livre. Outras possibilidades de apoio para este projeto incluem:

- parceria com o Centro de Competência em Software Livre (CCSL/USP) no âmbito do projeto “Painel de Gestão Acadêmica” (desenvolvido no ICMC/USP), que possui recursos de visualização de redes de colaboração e de genealogia acadêmica a partir do Lattes.
- Perspectivas de colaboração em HCI com Stephen Payne (Universidade de Bath, Inglaterra) e em analítica visual de texto e de rede com Evangelos E. Milios (Universidade de Dalhousie, Canadá).

## Referências

- [1] M. Newman, “Networks: an introduction. 2010,” *United States: Oxford University Press Inc., New York*, pp. 1–2.
- [2] P. Gundecha and H. Liu, “Mining social media: a brief introduction,” in *New Directions in Informatics, Optimization, Logistics, and Production*, pp. 1–17, Informs, 2012.
- [3] R. Fabbri, *Topological stability and textual differentiation in human interaction networks: statistical analysis, visualization and linked data*. PhD thesis, São Carlos Institute of Physics, University of São Paulo (IFSC/USP), 2017. <https://github.com/ttm/thesis/raw/master/tese-rfabbri.pdf>.
- [4] D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, *Mastering the information age solving problems with visual analytics*. Eurographics Association, 2010.
- [5] B. Shneiderman, “The eyes have it: A task by data type taxonomy for information visualizations,” in *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pp. 336–343, IEEE, 1996.
- [6] D. A. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler, “Challenges in visual data analysis,” in *Information Visualization, 2006. IV 2006. Tenth International Conference on*, pp. 9–16, IEEE, 2006.
- [7] E. Gröller, “Visual analytics - introduction.” University Lecture, 2015.

- [8] I. H. Witten, “Text mining.,” 2004.
- [9] F. Beck, M. Burch, S. Diehl, and D. Weiskopf, “A taxonomy and survey of dynamic graph visualization,” in *Computer Graphics Forum*, Wiley Online Library, 2016.
- [10] T. Berners-Lee, “Design issues: Linked data,” 2006.
- [11] “Nexos: Teoria crítica e pesquisa interdisciplinar.” <http://dgp.cnpq.br/dgp/espelhogrupo/5624425913774111>.
- [12] R. Fabbri, R. Fabbri, D. C. Antunes, M. M. Pisani, and O. N. Oliveira Jr, “Temporal stability in human interaction networks,” *arXiv preprint arXiv:1310.7769*, 2013.
- [13] R. Motta, R. Minghim, A. de Andrade Lopes, and M. C. F. Oliveira, “Graph-based measures to assist user assessment of multidimensional projections,” *Neurocomputing*, vol. 150, pp. 583–598, 2015.
- [14] R. Etemadpour, R. Motta, J. G. de Souza Paiva, R. Minghim, M. C. F. de Oliveira, and L. Linsen, “Perception-based evaluation of projection methods for multidimensional data visualization,” *IEEE transactions on visualization and computer graphics*, vol. 21, no. 1, pp. 81–94, 2015.
- [15] R. Fabbri, R. B. de Luna, R. A. P. Martins, *et al.*, “Social participation ontology: community documentation, enhancements and use examples,” *arXiv preprint arXiv:1501.02662*, 2015.
- [16] R. Fabbri, “A connective differentiation of textual production in interaction networks,” 2013. <http://arxiv.org/abs/1412.7309>.
- [17] R. Fabbri and O. N. Oliveira Junior, “Linked open social data for scientific benchmarking.” <https://github.com/ttm/linkedOpenSocialData/raw/master/paper.pdf>, 2016.
- [18] R. Fabbri, “Versinus: a visualization method for graphs in evolution,” *arXiv preprint arXiv:1412.7311*, 2013. <http://arxiv.org/abs/1412.7311>.
- [19] A. J. Soto, A. Mohammad, A. Albert, A. Islam, E. Milios, M. Doyle, R. Minghim, and M. C. Ferreira de Oliveira, “Similarity-based support for text reuse in technical writing,” in *Proceedings of the 2015 ACM Symposium on Document Engineering*, pp. 97–106, ACM, 2015.



- [20] D. M. Eler, F. V. Paulovich, M. C. F. de Oliveira, and R. Minghim, “Coordinated and multiple views for visualizing text collections,” in *Information Visualisation, 2008. IV’08. 12th International Conference*, pp. 246–251, IEEE, 2008.
- [21] F. San Roman, R. Pinho, R. Minghim, and M. de Oliveira, “A study on the role of similarity measures in visual text analytics,” in *Proceedings VISIGRAPP-8th. International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pp. 429–438, 2013.
- [22] R. Makki, E. Carvalho, A. J. Soto, S. Brooks, M. C. F. De Oliveira, E. E. Milios, and R. Minghim, “Atr-vis: Visual and interactive information retrieval for parliamentary discussions in twitter,” *ACM Transactions on Knowledge Discovery from Data (TKDD)*, Accepted January, vol. 2017, 2017.
- [23] M. D. S. d. S. Dias, M. R. Mansour, L. d. M. Romanetto, M. C. F. d. Oliveira, L. G. Nonato, *et al.*, “Vizlattes: a tool for relevance analysis from scientific co-authorship networks,” in *Conference on Graphics, Patterns and Images, XXVIII; Workshop on Visual Analytics, Information Visualization and Scientific Visualization, VI*, Sociedade Brasileira de Computação-SBC, 2015.
- [24] A. C. Telea, *Data visualization: principles and practice*. CRC Press, 2014.
- [25] W. Cleveland, *Visualizing data*. Hobard Press, 1993.
- [26] S. van den Elzen, D. Holten, J. Blaas, and J. J. van Wijk, “Reordering massive sequence views: Enabling temporal and structural analysis of dynamic networks,” in *Visualization Symposium (PacificVis), 2013 IEEE Pacific*, pp. 33–40, IEEE, 2013.
- [27] S. van den Elzen, D. Holten, J. Blaas, and J. J. van Wijk, “Reordering massive sequence views: Enabling temporal and structural analysis of dynamic networks,” in *Visualization Symposium (PacificVis), 2013 IEEE Pacific*, pp. 33–40, IEEE, 2013.
- [28] D. Koop, J. Freire, and C. T. Silva, “Visual summaries for graph collections,” in *Visualization Symposium (PacificVis), 2013 IEEE Pacific*, pp. 57–64, IEEE, 2013.
- [29] V. Gupta, G. S. Lehal, *et al.*, “A survey of text mining techniques and applications,” *Journal of emerging technologies in web intelligence*, vol. 1, no. 1, pp. 60–76, 2009.

- [30] F. V. Ordenes, B. Theodoulidis, J. Burton, T. Gruber, and M. Zaki, “Analyzing customer experience feedback using text mining: A linguistics-based approach,” *Journal of Service Research*, vol. 17, no. 3, pp. 278–295, 2014.
- [31] J. J. Thomas, *Illuminating the path:[the research and development agenda for visual analytics]*. IEEE Computer Society, 2005.
- [32] A. J. Bremner, S. Caparos, J. Davidoff, J. de Fockert, K. J. Linnell, and C. Spence, “Bouba and kiki in namibia? a remote culture make similar shape–sound matches, but different shape–taste matches to westerners,” *Cognition*, vol. 126, no. 2, pp. 165–172, 2013.
- [33] H. Sanoff, *Visual Research Methods in Design (Routledge Revivals)*. Routledge, 2016.
- [34] K. Koffka, *Principles of Gestalt psychology*, vol. 44. Routledge, 2013.
- [35] S. E. Palmer, K. B. Schloss, and J. Sammartino, “Visual aesthetics and human preference,” *Annual review of psychology*, vol. 64, pp. 77–107, 2013.
- [36] J. P. Mena-Chalco and C. Júnior, “Prospecção de dados acadêmicos de currículos lattes através de scriptlattes,” *Bibliometria e Cientometria: reflexões teóricas e interfaces. São Carlos: Pedro & João*, 2013.
- [37] U. Brandes, “A faster algorithm for betweenness centrality,” *Journal of mathematical sociology*, vol. 25, no. 2, pp. 163–177, 2001.
- [38] T. Pedersen, S. Patwardhan, and J. Michelizzi, “Wordnet:: Similarity: measuring the relatedness of concepts,” in *Demonstration papers at HLT-NAACL 2004*, pp. 38–41, Association for Computational Linguistics, 2004.
- [39] M. Krzywinski, I. Birol, S. J. Jones, and M. A. Marra, “Hive plots approach to visualizing networks,” *Briefings in bioinformatics*, vol. 13, no. 5, pp. 627–644, 2012.
- [40] “Data.world - datasets for analysis & download.” <https://data.world>,.
- [41] R. Fabbri, *Música no áudio digital: descrição psicofísica e caixa de ferramentas*. PhD thesis, Universidade de São Paulo.
- [42] R. Fabbri and R. S. Mendes, “Compressão de áudio via wavelets, aproximações polinomiais e permutações,” *ENCONTRO DOS ALUNOS E DOCENTES DO DEPARTAMENTO DE COMPUTAÇÃO E AUTOMAÇÃO INDUSTRIAL (EADCA)*, vol. 2, pp. 155–158, 2009.

- [43] R. Fabbri and A. MAIA JUNIOR, “Applications of group theory on granular synthesis,” *SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO MUSICAL*, vol. 11, 2007.
- [44] R. Fabbri, *A distance metric between histograms derived from the Kolmogorov-Smirnov test statistic: specification, measures reference and example uses.* <https://github.com/ttm/kolmogorov-smirnov/raw/master/paper.pdf>.
- [45] “Currículo lattes de renato fabbri.” <http://lattes.cnpq.br/1840472218825589>.
- [46] R. Fabbri, V. V. d. S. Junior, A. C. S. Pessotti, and D. C. Corrêa, “Psychophysics of musical elements in the discrete-time representation of sound,” *arXiv preprint arXiv:1412.6853*, 2014.