

# Um Método para Apoiar a Identificação de Interesses entre Participantes de Ecossistemas de Startups

Alternative Title: A Method to Support the Identification of Interests in Startup Ecosystems

Rafael E. L. Escalfoni  
CEFET-RJ  
Nova Friburgo, RJ  
rafael.escalfoni@cefet-rj.br

Tiago C. França  
DECOMP - UFRJ  
Seropédica, RJ  
tiago.franca@ufrj.br

Marcelo A. S. Irineu  
CEFET-RJ  
Rio de Janeiro, RJ  
marcelo.irineu@cefet-rj.br

Adriana S. Vivacqua  
PPGI - UFRJ  
Rio de Janeiro, RJ  
avivacqua@dcc.ufrj.br

Jonice Oliveira  
PPGI - UFRJ  
Rio de Janeiro, RJ  
jonice@dcc.ufrj.br

## ABSTRACT

Startups ecosystem is a creative environment to aim promoting entrepreneurship and the birth of innovative business. In this complex social network, the activities derive from the relationship of interdependence and co-evolution between actors with different purposes, who share resources to achieve common goals. The dynamic of this environment requires understanding the interests and capacities of each participant as a way to identify convergences and possible partnerships that can help in the development of new businesses. This paper presents a method to assist the identification interests and vocations from relationships extracted from interactions in social network tools in startups ecosystems. To verify applicability, an empirical study was conducted from a conversation group by Whatsapp held by managers of technology business incubators from Brazilian federal institutes of education and research.

## CCS CONCEPTS

• **Information systems** → **Expert systems**; *Social networks*;

## KEYWORDS

Startup Ecosystem; business networking; information retrieval; Social Network Analysis

## ACM Reference Format:

Rafael E. L. Escalfoni, Tiago C. França, Marcelo A. S. Irineu, Adriana S. Vivacqua, and Jonice Oliveira. 2018. Um Método para Apoiar a Identificação de Interesses entre Participantes de Ecossistemas de Startups: Alternative Title: A Method to Support the Identification of Interests in Startup Ecosystems. In *Proceedings of SBSI 2018 (SBSI'18)*, Jennifer B. Sartor, Theo D'Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, Article 4, 8 pages. [https://doi.org/10.475/123\\_4](https://doi.org/10.475/123_4)

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SBSI'18, June 2018, Caxias do Sul, Rio Grande do Sul, Brazil

© 2018 Copyright held by the owner/author(s).

ACM ISBN 123-4567-24-567/08/06.

[https://doi.org/10.475/123\\_4](https://doi.org/10.475/123_4)

## 1 INTRODUÇÃO

Os ecossistemas de *startups* são ambientes criativos que desempenham um importante papel na geração de inovação e desenvolvimento econômico [9]. Nesses espaços, os empreendedores procuram validar ideias inovadoras em um curto intervalo de tempo e transformá-las em negócios disruptivos com baixos investimentos iniciais [7].

Nessa comunidade empreendedora, cada membro possui um objetivo específico. Segundo Dedehayir *et al.* [10], os empreendedores são essenciais para o surgimento de novos negócios, as organizações de apoio, tais como incubadoras e aceleradoras, desempenham o papel de facilitadores no processo de desenvolvimento de novos empreendimentos, enquanto os grupos de investidores dão o aporte financeiro necessário para o crescimento em escala. Essa natureza diversa dos participantes contribui com diferentes perspectivas e conduz ao enriquecimento das ideias que irão resultar em novos produtos e modelos de negócios [26].

Em um ambiente consolidado, há um grande número de participantes, aumentando as possibilidades de parcerias. A pluralidade pode influenciar positivamente no desenvolvimento de negócios [27]. Porém, a identificação de parceiros adequados para projetos vai se tornando uma tarefa mais difícil, pois é necessário conjugar os interesses mútuos [16]. Neste contexto, é importante facilitar a formação de parcerias por meio de recomendações baseadas nos interesses dessas redes tipicamente dinâmica, complexas e de pessoas que se relacionam utilizando mídias para comunicação textual. Essas recomendações passam pela identificação dos assuntos discutidos, interesses dos participantes e possível reconhecimento de especialistas para recomendação de parcerias na rede. Há alguns estudos acerca do assunto [1, 9], contudo faltam métodos e mecanismos de apoio a esta tarefa [12, 14].

Neste sentido, fizemos um primeiro estudo que propõe um método para auxiliar na identificação de interesses, especialistas e influenciadores a partir das relações extraídas de interações em mídias sociais online formada por atores de um ecossistema empreendedor. Em seguida, é apresentado um experimento que foi realizado para fazer uma avaliação inicial da aplicabilidade da abordagem, a partir de uma amostragem com quatro meses de conversas coletadas de um grupo de *Whatsapp* mantido por gerentes de incubadoras tecnológicas de institutos federais. O grupo de conversas tem sido

utilizado como uma importante ferramenta para a comunicação diária.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta os conceitos que fundamentam este trabalho. A Seção 3 detalha os passos do método proposto, suas fases e métricas aplicadas. A Seção 4 traz uma discussão sobre o planejamento e execução do experimento realizado. Na Seção 5 são apresentadas as conclusões da pesquisa e os trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção serão apresentados e discutidos os conceitos necessários para a compreensão do método proposto no presente trabalho. Na Seção 2.1, será apresentado o conceito de *Ecosistema de startups* e seus participantes. A Seção 2.2 define o conceito de análise de redes sociais e as métricas aplicadas no estudo. Em seguida, a Seção 2.3 apresenta alguns trabalhos relacionados.

### 2.1 Ecossistemas de *Startups*

O modelo de *startups* tem sido tratado com grande expectativa pela eficiência em experimentar conceitos e verificar a viabilidade de produtos e negócios até que se tenha condições mínimas para operacionalizar uma empresa [7]. Seus passos compreendem o desenvolvimento de ideias e um projeto de organização que visam identificar um modelo de negócio que possa ser expandido [18]. Essa dinâmica reduz o custo e o tempo para o lançamento de produtos, permitindo interromper o ciclo rapidamente, caso os resultados parciais não sejam satisfatórios, o que possibilita lidar com maiores riscos e incertezas [9].

O sucesso das *startups* é influenciado por uma série de fatores. Dentre os quais, o meio onde as iniciativas se desenvolvem: é imprescindível dispor de organizações de apoio que ofereçam infraestrutura, serviços administrativos e redes de contatos para facilitar o acesso a prestadores de serviços, recursos tecnológicos e financiamento [19]. A participação neste ambiente de negócios requer aceitação através da convergência de interesses que levarão ao crescimento mútuo [5]. Conforme percebido por Moore [20], tais comunidades manifestam-se como ecossistemas biológicos, onde as atividades de negócio se traduzem em relações de interdependência e coevolução.

Um ecossistema de startups pode ser definido como um conjunto de fatores que promovem o espírito empreendedor, acompanham e apoiam o processo de desenvolvimento de *startups* e desempenham um papel no desenvolvimento do empreendedorismo [28]. Esta complexa rede social de negócios é composta por atores com diferentes propósitos. As conexões formadas são responsáveis pela troca de experiências entre empreendedores e o acesso a recursos complementares para o desenvolvimento de produtos, tais como tecnologias e linhas de financiamento [12]. Mota *et al.* [21] definem os seguintes atores, conforme ilustrado na Figura 1:

- Os **Empreendedores** são as pessoas que buscam oportunidades de negócio para abrir novas empresas. Torres e Souza [9] enfatizam que em países em desenvolvimento muitas pessoas empreendem por falta de bons empregos, são os chamados empreendedores por necessidade. Em contrapartida existem aqueles que estão em busca de novos desafios, os chamados empreendedores seriais.



Figura 1: Atores de um Ecosistema de Startups[21].

- As **Organizações de Apoio** atuam no desenvolvimento, suporte e encorajamento de ações empreendedoras. Elas trabalham para assegurar que os empreendedores disponham de todos os recursos necessários para aumentar a chance de êxito das propostas [2, 14, 26]. Para isto, analisam as necessidades de cada projeto e lançam mão de suas redes sociais para fazerem a ponte entre os participantes do ecossistema, assegurando a convergência de interesses e o crescimento mútuo [5]. Incubadoras, aceleradoras e parques tecnológicos se enquadram neste perfil de participantes.
- O **Governo** representa o sistema político que controla e regulamenta a região. Suas ações podem interferir diretamente no ecossistema. Através de subsídios, incentivos fiscais e políticas públicas, pode tornar a comunidade mais atraente e robusta [2, 8].
- **Provedores de Serviços** são as organizações que executam as atividades de suporte das *startups*, tais como escritórios de contabilidade, advogados, consultores, etc.
- **Grandes Empresas** buscam talentos e incentivam o surgimento de novos parceiros. Segundo Águeda [2], as empresas consolidadas buscam na regionalidades uma forma de procurar soluções criativas. A colaboração com as *startups* auxiliam na identificação de novos segmentos de mercado e ajuda a fomentar a inovação [17]. Podem oferecer recursos e suporte financeiro, além de aconselhamento.
- **Instituições de Ensino e Pesquisa** são responsáveis pela capacitação e desenvolvimento de novas tecnologias. Apesar de possuírem um grande potencial no campo das descobertas científicas, não são vocacionadas para o lançamento de produtos comerciais. Atualmente, o empreendedorismo tem ganhado força nas universidades. Com a percepção de que o empreendedorismo poderia ser ensinado e que o apoio das instituições contribuiria para o surgimento de negócios sustentáveis, as grandes universidades passaram a atuar cada vez mais no tema [25].

Torres e Souza [9] destacam a natureza complexa das relações em rede, como sendo resultado de uma combinação única em cada região fruto da cultura de apoio e da participação de empresários,

governos, investidores e universidades. Compreender a natureza das relações entre estes diferentes atores é crítico para o sucesso do ecossistema.

## 2.2 Análise de Redes Sociais

O entendimento das relações em um ecossistema empreendedor pode ser facilitado pelo uso da teoria de redes sociais [22]. Uma rede social pode ser caracterizada como um conjunto de atores que se relacionam com outros atores [13]. O relacionamento pode ser de amizade, trabalho, entre outros. O relacionamentos são dinâmicos: novos amigos surgem, muda-se de emprego, amizades se desfazem, etc. O fluxo de conhecimento da rede social citado por Neumeyer e Santos [22] passou, nos últimos anos, a fluir em sistemas distribuídos chamados mídias sociais. Nas mídias sociais, novas interações entre pessoas podem ocorrer motivadas por interesses comuns ainda que tais pessoas não se conheçam pessoalmente.

O estudo das propriedades de redes pode revelar padrões de interações entre pessoas e organizações. É comum representar as redes como sendo grafos: os atores são os nós e as interações as arestas dos grafos. Nós e arestas podem receber pesos diferenciados para representar a quantidade de conexões do nó ou frequência de interações. Dessa forma, é possível representar diferentes características de uma rede social. Por exemplo, os nós e arestas podem ser representados com ou sem peso considerando-se apenas as ligações entre os nós, ou com pesos. As arestas também podem utilizar pesos diferentes para indicar intensidade, número de ocorrências ou probabilidade das relações.

Considerando uma mídia social de conversas entre usuários, como a utilizada no estudo de caso deste trabalho, pode-se representar a rede de tópicos de discussão e as mensagens de respostas como os nós da rede. As respostas são ligadas ao tópico formando a rede. Outra rede pode ser formada considerando-se os autores e postagens como sendo nós de uma rede e autoria como aresta. Nesse caso, a rede pode ser modelada, por exemplo, como um grafo bipartido, com autores e postagens como dois conjuntos disjuntos de nós.

A análise de redes sociais (ARS) é o mapeamento e medição das relações entre pessoas, grupos, organizações e outras entidades, que permite traçar um diagnóstico gerencial. A ARS pode auxiliar na análise do nível de coordenação de parcerias, a intensidade das interações, o surgimento de comunidades, o nível de conectividade e a relevância de participantes [3].

Das métricas utilizadas para analisar redes sociais, utilizamos: o grau do nó, densidade e o grau de intermediação. O grau de um nó em uma rede social é o número de arestas conectadas a ele. Nos casos de redes direcionadas existem medidas de graus de entrada e saída, onde o grau de entrada corresponde ao número de arestas que chegam ao nó e o grau de saída se refere às arestas que saem dele [13].

A densidade de um nó, ou coeficiente de agrupamento, é a razão entre o número de vizinhos do nó e o número de vizinhos possíveis. Esta medida indica o quão bem conectado um nó está na rede [13]. Por exemplo, em um grafo de tópicos de discussão, um assunto com densidade alta indica grande interesse da comunidade. Em redes de autores e mensagens, pode indicar pessoas influentes e os interesses individuais [29]. A densidade da rede é a razão entre o número total

de relacionamentos da rede e o número total de relacionamentos possíveis. Esta métrica indica o quanto uma rede está perto de estar completa. No caso das redes de discussão, mede o nível de troca de informações ou coordenação [22]. Se a densidade do grafo for baixo, isto pode indicar falta de foco nas discussões.

O grau de intermediação representa o número de caminhos mínimos que passam por um nó. Um nó que possua um alto número de caminhos mínimos é importante para a rede. Em uma rede de parceiros de um ecossistema de *startups*, um alto grau de intermediação representa a importância do ator para a comunidade [22].

## 2.3 Trabalhos Relacionados

O aprimoramento das relações de interação em redes sociais vem sendo analisado em diferentes contextos. Na área acadêmica, há uma série de trabalhos buscando ampliar a colaboração e mapear competências de comunidades [11, 24]. Através de análises na plataforma de pesquisadores Lattes<sup>1</sup>, Digiampietri *et al.* [11] propuseram um sistema para predição de novas relações de coautoria, enquanto Oliveira e Paula [24] mapearam a integração entre pesquisadores segundo sua distribuição geográfica e os temas abordados.

A qualidade das interações e a confiança entre os membros de uma comunidade também tem sido abordado recentemente [1, 4, 23]. A identificação da reputação em redes sociais foi analisada por Neves e Ferreira [23] através de um método não supervisionado. O trabalho apresenta uma maneira de agrupar assuntos semelhantes e definir a relevância dos tópicos em uma base de *tweets*.

As relações de parcerias de negócio também já foram estudadas. No trabalho realizado por Acosta *et al.* [1], foram definidas métricas para avaliar as relações de parceria. Contudo, a análise depende da entrada de dados em uma plataforma específica para colaboração. O foco principal do estudo foi desenvolver um método para auxiliar na identificação de interesses, influenciadores e especialistas em grupos de discussão, independente de plataforma.

## 3 MÉTODO PROPOSTO

Este trabalho apresenta uma abordagem para identificar interesses em ecossistemas de *startups* a partir de interações por trocas de mensagens entre os participantes de canais de comunicação. Os assuntos discutidos são analisados para definir os mais discutidos, a influência e o interesse de cada pessoa e a repercussão das postagens na rede. Para isso, o processo foi dividido em quatro etapas (ilustradas na Figura 2): (1) Extração e anotação de mensagens; (2) Classificação de tópicos de discussão; (3) Montagem da rede; e (4) Análise de métricas.



Figura 2: Etapas do Método Proposto.

### 3.1 Etapas do Método Proposto

A primeira etapa corresponde a extração da base de dados e a escolha da abordagem para armazená-los de forma que ele possa ser

<sup>1</sup><http://lattes.cnpq.br/>

consultados, avaliados e pré-processados. Por exemplo, neste trabalho a base de dados de mensagens de comunicação no *WhatsApp* foi salva em um banco de dados relacional. Cada mensagem foi incluída como uma tupla do banco.

As mensagens são então anotadas como sendo tópicos de discussão - que são postagens que iniciam uma discussão sobre um assunto de interesse - ou como sendo resposta a um tópico. Um assunto é um conjunto de tópicos de discussão considerados similares. Neste trabalho, cada tópico de discussão é associado a um único assunto.

As mensagens de resposta são associadas ao tópico de discussão correspondente. Além disso, as mensagens de resposta também são marcadas de acordo com seu posicionamento acerca do tópico. O posicionamento pode ser positivo, negativo ou neutro. Uma mensagem pode ser considerada positiva quando se percebe concordância, apoio ou agradecimento associado ao tópico original. As respostas negativas trazem argumentos contrários ao tópico. Finalmente, as respostas neutras tratam de questionamentos ou qualquer outro conteúdo que não pode ser considerado positivo ou negativo.

Durante a segunda etapa, os tópicos de discussão são classificados através da plataforma de processamento de linguagem natural IBM Watson [15]. Para aumentar a confiabilidade dos resultados, a categorização resultante do algoritmo deve ser verificada por pares de pesquisadores. Por exemplo, os pesquisadores avaliam o tópico e verificam a qual assunto está relacionado. Isso acontece com base em uma lista de assuntos previamente definidos pelos pesquisadores a partir de uma análise de mensagens. Essa foi a abordagem adotada neste trabalho.

Na terceira etapa são modeladas as redes para se alcançar o objetivo final: identificar interesses, interessados e especialistas com base em mensagens de grupo de discussão. As redes são representadas por grafos nos quais dados de postagens e autores representam os nós do grafo. Utilizando esses nós, dois grafos foram definidos: o primeiro grafo com autores e assuntos representados por nós e arestas ligando os autores aos assuntos postados; e o outro grafo com assuntos e respostas representados por nós e as arestas ligando as respostas aos assuntos correspondentes.

Com as redes formadas, são realizadas as análises na quarta etapa. Para isto, foram definidos conceitos e métricas associados ao grafo gerado. As métricas de ARS foram utilizadas em algumas análises de forma adaptada para se conhecer as características básicas da topologia das redes. Considerando as redes formadas, as relações entre postagens e assuntos buscou-se obter as seguintes informações: repercussão do tópico, relevância de um assunto, interesse de um participante sobre um assunto e grau de influência do participante sobre um assunto e os especialistas.

## 3.2 Métricas Adotadas

Definimos a seguir os indicadores utilizados para extrair as informações pertinentes ao trabalho.

**3.2.1 Repercussão de um Assunto.** Representada na Equação 1, a repercussão de um assunto permite verificar a frequência de reações dos participantes para aquele tema. Uma repercussão alta pode indicar se tratar de um tópico polêmico ou de grande interesse de toda a comunidade. Definimos o cálculo da repercussão de um assunto  $x$  ( $a_x$ ) como o grau de entrada (*indegree*) do nó  $a$ .

$$\text{repercussão}_x = a_x \text{indegree} \quad (1)$$

**3.2.2 Relevância de um Assunto.** É uma medida relativa que auxilia na identificação dos temas mais comuns no fórum. Em mídias sociais mais sofisticadas, a determinação de temas auxiliaria na criação de canais de discussão específicos, evitando a dispersão de conversas. A relevância de um assunto  $x$  foi definida como a razão entre o grau de entrada dos tópicos de assunto  $x$ ,  $a_x$  pelo somatório de graus de entrada de todos os assunto identificados ( $a_i$ ), conforme apresentado na Equação 2.

$$\text{relevância}_x = \frac{a_x \text{indegree}}{\sum_{i=1}^n a_i \text{indegree}} \quad (2)$$

**3.2.3 Nível de Interesse.** É obtido pela identificação dos membros mais atuantes em um determinado assunto. O interesse é uma informação importante, pois ela identifica os membros que podem colaborar uns com os outros para realizar tarefas relacionadas.

A Equação 3 representa a fórmula construída para calcular o nível de interesse dos membros da rede por assunto abordado. O nível de interesse em um assunto  $x$  é calculado somando o número de ocorrências onde o membro  $y$  iniciou uma discussão sobre o tema com a soma de respostas desse usuário para o mesmo assunto. Atribuímos um peso  $\alpha$  para as vezes em que o membro inicia um tópico, e um peso  $\beta$  para as respostas do usuário. Tais parâmetros foram incluídos para possibilitar a diferenciação das ações. No exemplo do trabalho, atribuímos  $\alpha = 2$  e  $\beta = 1$ .

$$\text{interesse}_{x,y} = \alpha * t_{x,y} + \beta * r_{x,y} \quad (3)$$

onde:

$$\begin{cases} t_{x,y}, & \text{número de conversas sobre } x \text{ iniciadas por } y \\ r_{x,y}, & \text{número de respostas sobre } x \text{ feitas por } y \end{cases}$$

**3.2.4 Grau de Influência.** A influência na rede pode ser definida como o efeito que a ação de um indivíduo pode causar em seus pares, levando a agirem de forma similar [4]. Neste trabalho, a influência de um usuário na rede é analisada sobre assuntos iniciados por ele. A modelagem adotada permite que seja identificada a origem de uma discussão e as respostas endereçadas a ela. Ou seja, identifica-se os assuntos e então são formados *clusters* com postagens de resposta apontando para a mensagem inicial de um dado assunto.

A relação de usuários que respondem a um assunto pode variar. Por exemplo, um assunto  $a_1$  publicado em um grupo com 10 usuários pode obter 100 respostas, 95 publicadas por dois usuários diferentes do autor da mensagem que define o assunto. As outras 5 são de autoria do próprio usuário que iniciou o assunto. A fim de capturar essa relação de quantidade de usuários que respondem e volume de respostas, definimos dois parâmetros: mobilização e intensidade.

A mobilização ( $M$ ) é uma métrica relativa, utilizada para identificar quantos participantes do grupo se sensibilizaram com a mensagem inicial. Foi definida como o número de usuários afetados (conjunto  $U'$ ) dividido pelo total de participantes da rede (conjunto  $U$ ), excetuando o próprio participante. A intensidade foi elaborada visando analisar o volume de respostas obtidas dentre todo o grupo,

ou seja, a intensidade ( $T$ ) mede a razão entre respostas ( $R$ ) e número de colegas na rede ( $U - 1$ ).

Desta forma, a fórmula para calcular o grau de influência acerca de um assunto  $x$  por um usuário  $y$ , relacionando  $M$  e  $T$  é representada na Equação 4.

$$\text{influência}_{x,y} = \frac{(M + T)}{2} \quad (4)$$

$$\text{Onde: } M = \frac{U'}{U - 1} \quad \text{e} \quad T = \frac{R}{U - 1}$$

É importante destacar que a influência atribuída a um usuário está relacionada apenas a sua influência naquela rede, visto que mobilização e intensidade de respostas podem ter valores diferentes dependendo da quantidade de usuários. Ou seja, rede com muitas pessoas proporcionará valores menores do que redes com poucos usuários. Portanto, a influência de um usuário em um grupo não tem validade semântica para representar sua influência como “ser” não podendo um grau de influência de um grupo ter significado em outros contexto que não seja aquele grupo.

**3.2.5 Expertise.** Está relacionada ao prestígio e reputação de um indivíduo em relação a um assunto. Em seu trabalho de detecção de especialistas em comunidades online, Zhang *et al.* [31] afirmam que pessoas que recebem muitas respostas positivas são consideradas de prestígio, trazendo indícios de maior afinidade e importância do indivíduo naquele assunto. As respostas podem ser obtidas quando um usuário envia uma mensagem para o grupo se manifestando com relação a um assunto ou quando ele responde a dúvidas de um colega e é apoiado por membros do grupo.

Para calcular a *expertise* de um usuário  $y$  em um determinado assunto  $x$ , identificamos o apoio recebido às suas mensagens no grupo através da razão entre a quantidade de mensagens positivas recebidas após a publicação de um tópico de discussão pelo total de mensagens direcionadas àquele assunto.

Suas contribuições para o grupo através de suas respostas para questões levantadas também são importantes indícios de sua competência no assunto. Então calculamos a razão das respostas do usuário e todas as respostas direcionadas para o assunto.

As respostas são classificadas como positiva - quando demonstram algum grau de apoio aos argumentos do autor, ou como não-positiva - perguntas, mensagens neutras, mensagens negativas, etc.

A fórmula de cálculo da *expertise* de um usuário  $y$  em um assunto  $x$  foi definida como a média aritmética entre a razão das mensagens positivas e o total de mensagens e a razão entre suas respostas e o total de respostas do grupo acerca do assunto, como demonstrado na Equação 5.

$$\text{expertise}_{x,y} = \frac{\frac{P_{x,y}}{R_x} + \frac{A_{x,y}}{A_x}}{2} \quad (5)$$

Onde:

$$\left\{ \begin{array}{l} P_{x,y} \quad \text{é o número de mensagens positivas endereçadas ao assunto } x \text{ recebidas por um autor } y. \\ R_x \quad \text{é o número total de mensagens (positivas e não-positivas) para o assunto } x. \\ A_{x,y} \quad \text{é o número de respostas do autor } y \text{ para o assunto } x. \\ A_x \quad \text{é o número total de respostas endereçadas ao assunto } x. \end{array} \right.$$

## 4 RESULTADOS OBTIDOS

Nesta seção, apresentamos um estudo empírico realizado com o objetivo de verificar indícios da aplicabilidade do método proposto. Para isto, foi planejada uma pesquisa com dados coletados a partir de um grupo de conversas da ferramenta *WhatsApp*. O grupo é mantido por gerentes de incubadoras de institutos federais e representa um importante canal de comunicação entre os participantes. Conforme relatos de seus usuários, a ferramenta exerce um papel marcante de identificação do grupo como uma rede, por onde há troca de experiências, alinhamento de estratégias e procedimentos comuns, respeitando as regionalidades. O grupo dos gerentes possui 35 integrantes, de instituições de 17 estados espalhados por todas as regiões do país.

As incubadoras tecnológicas desempenham um importante papel na articulação de empreendimentos, oferecendo espaço físico, oportunidades para transferência tecnológica, consultorias e contato com investidores e entidades governamentais [19]. O experimento buscou identificar os interesses e protagonistas do grupo de conversas aplicando as métricas propostas.

### 4.1 Execução do Método

A realização da primeira etapa do método consistiu na coleta e tratamento dos dados. A amostra disponibilizada tinha quase quatro meses de conversas contendo 1666 postagens, sendo a primeira mensagem feita no dia 22/02/2017 e a última no dia 12/06/2017. Foram identificados 267 tópicos de discussão e 946 respostas. Apenas foram consideradas as mensagens de texto. Como o *WhatsApp* é uma ferramenta de comunicação informal, muitas mensagens estavam fragmentadas em mais de uma postagem. Foram encontrados 444 trechos complementares de mensagem, que foram incorporadas nas mensagens originais. Algumas postagens foram removidas por se tratarem de ruídos - saudações, imagens, áudios e etc.

A categorização das mensagens foi realizada inicialmente com a plataforma IBM Watson. Para verificar a taxa de acerto da ferramenta, dois pesquisadores verificaram as respostas uma a uma. A taxa de concordância com a classificação automática foi baixa: em torno de 32%. Isto certamente ocorreu por conta da natureza das mensagens de *WhatsApp* - normalmente curtas e informais. Desta forma, optou-se por fazer anotações manuais nas mensagens feitas por dois pesquisadores para reduzir a chance de erros de interpretação. Foram identificadas 31 classes e 118 subclasses de mensagens, totalizando 158 possibilidades de anotação. Os assuntos mais recorrentes no período analisado foram: (a1) Contatos entre gestores, (a2) Parcerias / Anprotec, (a3) Edital - seleção de projetos / Pré-incubação, (a4) Evento com incubadoras e (a5) Divulgação de evento com associação / Anprotec.

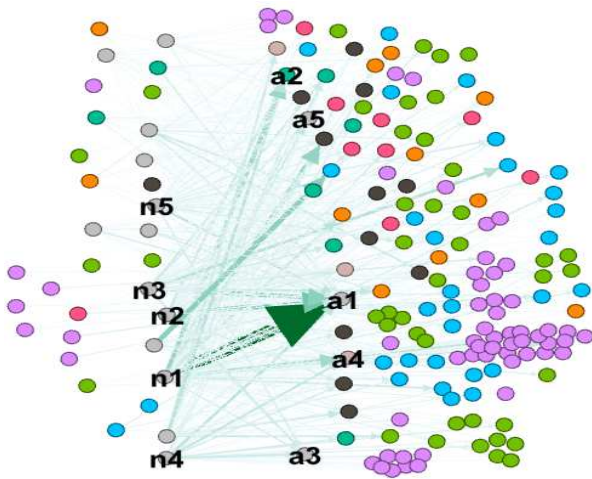


Figura 3: Grafo de Autores e Assuntos.

Durante a etapa de análise de redes de mensagens, foram montados dois grafos utilizando a ferramenta Gephi [6]. No primeiro grafo, apresentado na Figura 3, utilizamos um grafo bipartido, direcionado e ponderado. Os nós representam os usuários e assuntos e as arestas são as postagens dos usuários sobre o tema de destino. O grande número de classificações levou a um alto nível de dispersão no grafo de autores: de todos os assuntos iniciados, apenas 20 tem mais de 10 postagens, cerca de 50% das conversas do grupo. A partir desta rede, foi possível determinar as métricas de repercussão e relevância dos assuntos.

Das discussões mais frequentes no grupo, a que teve a maior repercussão (repercussão = 125), envolveu o *networking* dos participantes: para contactar alguém, intermediar alguma atividade, solicitar algum contato. Em seguida, estão as conversas acerca de parcerias com uma associação e o lançamento de editais de seleção de empreendedores. O somatório dos graus de entrada de todos os assuntos é 904, logo a *relevância*  $a_1 = 125/904 \approx 0,1383$ , conforme detalhado na Tabela 1.

Tabela 1: Métricas aplicadas às redes

Assunto	Repercussão	Relevância
a1	125	0,1383
a2	29	0,0321
a3	26	0,0288
a4	25	0,0277
a5	21	0,0232

Em seguida, foi construído um grafo direcionado para tratar do volume de respostas para os autores, que pode ser observado na Figura 4. Nesta rede, as mensagens de resposta e os assuntos correspondem aos nós e as arestas representam as relações entre postagens e assuntos, ou seja, respostas e reações a tópicos iniciados. Nota-se nesta rede a formação de *clusters* em torno dos assuntos discutidos. No centro da figura, o maior aglomerado corresponde às discussões acerca do assunto a1.

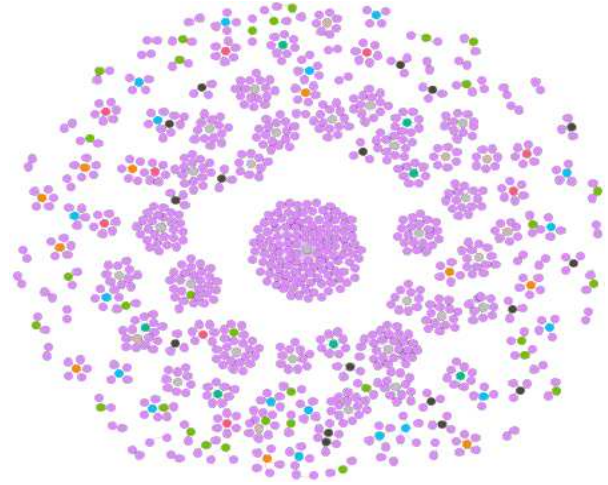


Figura 4: Grafo de Discussões por Assunto.

Optou-se por verificar o nível de interesse dos assuntos identificados como mais relevantes, com um peso  $\alpha = 2$  e  $\beta = 1$  para dar maior relevância aos tópicos postados. Também foi sobre estes assuntos que se buscou o grau de influência e a *expertise* de até cinco dos participantes mais interessados.

Para o assunto a1, o participante n1 iniciou 18 conversas ( $t_{a1, n1} = 18$ ) e publicou 14 respostas a colegas ( $r_{a1, n1} = 14$ ), logo  $interesse_{a1, n1} = 2 * 18 + 14 = 50$ . Para o cálculo da influência é preciso considerar há 35 participantes ativos no grupo ( $U = 35$ ) e que o número de colegas que se manifestaram quando n1 falou sobre a1 foi 13 ( $U' = 13$ ), que postaram 16 mensagens de resposta a tópicos de discussão iniciados por n1. Desta forma:

$$\left\{ \begin{array}{l} M = \frac{13}{35 - 1} \approx 0,3823 \\ T = \frac{16}{35 - 1} \approx 0,4706 \\ \text{influência}_{a1, n1} = \frac{0,3823 + 0,4706}{2} \approx 0,426 \end{array} \right.$$

Para o assunto a1, foram geradas um total de 100 mensagens, entre tópicos e respostas ( $R_{a1} = 100$ ). O participante n1 obteve 3 mensagens positivas ( $P_{a1, n1} = 3$ ). O autor n1 postou 14 respostas ao assunto a1 e o número total de respostas geradas pelo grupo foi de 52. Sendo assim:

$$\text{expertise}_{a1, n1} = \frac{3}{100} + \frac{14}{52} \approx 0,149$$

Os demais valores estão detalhados na Tabela 2.

Os resultados obtidos indicam que os participantes n1, n2, n3 e n4 são os mais influentes do grupo para os assuntos analisados. Conforme observado na rede, estes são os gestores mais experientes, com incubadoras mais consolidadas ou que ocupam cargos de coordenação em associações e outras entidades de apoio ao empreendedorismo. Também são os membros mais antigos no grupo de conversas. A nossa amostra abrange as mensagens desde o início do

Tabela 2: Métrica de Autores por Assuntos

Assunto	Nível de Interesse	Grau de Influência	Expertise
a1	n1 = 50	n1 ≈ 0,426	n1 ≈ 0,149
	n3 = 19	n4 ≈ 0,206	n2 ≈ 0,073
	n2 = 14	n2 ≈ 0,132	n3 ≈ 0,067
	n4 = 11	n16 ≈ 0,103	n4 ≈ 0,039
	n5 = 7	n23 ≈ 0,103	n5 ≈ 0,029
a2	n1 = 10	n5 ≈ 0,118	n1 ≈ 0,364
	n2 = 9	n6 ≈ 0,088	n2 ≈ 0,157
	n6 = 7	n14 ≈ 0,044	n3 ≈ 0,045
	n5 = 4	n1 ≈ 0,029	n4 ≈ 0,045
	n14 = 2	n2 ≈ 0,029	n6 ≈ 0,045
a3	n1 = 5	n2 ≈ 0,176	n2 ≈ 0,153
	n3 = 5	n11 ≈ 0,118	n3 ≈ 0,139
	n4 = 5	n7 ≈ 0,088	n1 ≈ 0,125
	n5 = 4	n1 ≈ 0,059	n4 ≈ 0,125
	n2 = 3	n4 ≈ 0,059	n5 ≈ 0,111
a4	n1 = 13	n5 ≈ 0,176	n2 ≈ 0,21
	n2 = 13	n3 ≈ 0,132	n1 ≈ 0,136
	n5 = 9	n14 ≈ 0,132	n5 ≈ 0,095
	n3 = 6	n1 ≈ 0,118	n3 ≈ 0,079
	n4 = 4	n9 ≈ 0,103	n9 ≈ 0,074
a5	n1 = 7	n2 ≈ 0,265	n2 ≈ 0,329
	n2 = 4	n1 ≈ 0,132	n3 ≈ 0,133
	n3 = 4		n1 ≈ 0,112
	n5 = 3		n5 = 0,1

grupo até quatro meses de conversas, período em que novos membros foram sendo adicionados. É possível que outros importantes atores não tenham tido a oportunidade de se manifestar.

## 4.2 Ameaças à Validade

Este experimento, assim como qualquer estudo, possui diversas limitações e ameaças que podem afetar a validade dos resultados. Apresentamos a seguir as ameaças identificadas, segundo as diretrizes propostas por Wohlin *et al.*[30].

- Validade de construção: o método definido depende do significado das mensagens para classificá-las e anotá-las de acordo com as relações percebidas. Para reduzir essa ameaça, inicialmente tentou-se classificar automaticamente as mensagens, utilizando o IBM Watson, porém a informalidade das conversas por meio de mensagem utilizando a ferramenta *WhatsApp* fez com que esta abordagem tivesse uma acurácia muito baixa. A saída para o desenvolvimento deste trabalho foi a anotação manual realizada por dois pesquisadores. Ajustes nas respostas foram sendo realizados continuamente ao longo do experimento. Apesar do risco de enviesar as anotações pela expectativa dos investigadores, houve uma alta taxa de convergência entre os pares.
- Validade interna: a ferramenta *WhatsApp* possibilita a utilização de figuras, fotos e outros recursos de mídia para a troca de mensagens. Como nossa abordagem pretende analisar

conteúdos textuais, algumas mensagens importantes podem ter sido descartadas.

- Validade externa: neste experimento não houve uma população estabelecida estatisticamente. Como o grupo é dinâmico, novos participantes foram sendo adicionados ao longo do tempo, dificultando uma representação precisa para as métricas definidas. É preciso executar outros experimentos com um universo amostral maior para ampliar os indícios de aplicabilidade do método.
- Validade de conclusão: este trabalho traz um experimento inicial para a identificação de assuntos, interesses e especialistas. Desta forma, as métricas precisam passar por mais testes para serem aprimoradas.

## 5 CONCLUSÕES E TRABALHOS FUTUROS

O modelo de *startup* tem sido tratado como uma importante ferramenta de experimentação e aprendizagem para a geração de inovações. Sua abordagem permite testar novas ideias com o mínimo de recursos possível [7]. O sucesso de uma *startup* é influenciado pelo ambiente no qual está inserido e das relações mantidas entre seus agentes [2, 5, 9].

Este ecossistema, onde as *startups* se desenvolvem, possui atores dos mais diversos. Esta pluralidade pode fazer surgir negócios mais inovadores [14, 27]. Contudo, encontrar os parceiros mais adequados pode não ser tão simples. É preciso compreender as necessidades de cada participante da rede, seus interesses e suas competências. O presente artigo apresenta uma método que utiliza canais de discussão de grupos para auxiliar na identificação de assuntos de interesse, especialistas, influenciadores.

Foi conduzido um estudo empírico para verificar a aplicabilidade do método. Para isto, utilizou-se conversas extraídas de um grupo de *WhatsApp* formado por gerentes de incubadoras tecnológicas de institutos federais distribuídos por todo país. Este canal de comunicação é amplamente utilizado para discutir, divulgar eventos, compartilhar materiais, tirar dúvidas, etc. Os resultados obtidos através do experimento não podem ser generalizados, mas trazem indícios de que método auxiliou na identificação dos assuntos discutidos e na compreensão da relevância dos temas, reconhecimento de especialistas, dos temas de interesse de cada participante e dos influenciadores.

Em trabalhos futuros, esperamos ampliar a compreensão acerca das relações de parcerias em ecossistemas de *startups* por meio de uma análise investigativa mais detalhada. Pretendemos expandir a aplicação do método em outros grupos, tais como aceleradoras, empreendedores e grupos de investidores. É necessário também, aprimorar as métricas heurísticas propostas neste artigo e desenvolver ferramentas que facilitem a execução do método. O objetivo é dar aos participantes do ambiente empreendedor uma visão mais ampla das possibilidades de parcerias e criar mecanismos de gestão de redes de cooperação.

## ACKNOWLEDGMENTS

Gostariamos de agradecer as Incubadora de Empresas Tecnológicas de Institutos Federais por todo apoio ao projeto. Ao CNPq, CAPES e FAPERJ. Ao CEFET/RJ pelo apoio através da portaria

nº 122 de 17/01/2018. À UFRRJ pelo apoio através do processo nº 23083.004483/2015-39.

## REFERÊNCIAS

- [1] J. Acosta, A. Chandra, L. Xenakis, D. Varda, M. Eros-Sarnyai, D. Eisenman, I. Gonzalez, J. Gutierrez, D. Glick, and S. Sprong. 2015. Partnerships for Recovery across the Sectors (PRACTIS) Toolkit. <http://www.rand.org/pubs/tools/TL188>
- [2] A. F. P. Águeda. 2016. *Interconnectivity between Ecosystem Builders and Investor Groups in European Startup Ecosystems*. Master's thesis. Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.
- [3] N. Albaloooshi, N. Mavridis, and N. Al-Qirimi. 2012. A Survey on Social Networks and Organization Development. In *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, 539–545.
- [4] A. Anagnostopoulos, R. Kumar, and M. Mahdian. 2008. Influence and Correlation in Social Networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 7–15.
- [5] C. Bandera and E. Thomas. 2017. Startup Incubators and the Role of Social Capital. In *IEEE Technology & Engineering Management Conference (TEMSCON)*. IEEE, 142–147.
- [6] M. Bastian, S. Heymann, and M. Jacomy. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. In *Third International ICWSM Conference*.
- [7] S. Blank and B. Dorf. 2012. *The Startup Owners Manual: The Step-by-step Guide for Building a Great Company*. K&S Ranch, Pescadero, USA.
- [8] D. Cukier, F. Kon, and T. S. Lyons. 2016. Software Startup Ecosystems Evolution: The New York City Case Study. In *2nd International Workshop on Software Startups, IEEE International Technology Management Conference (Trondheim)*. IEEE.
- [9] N. N. de J. Torres and C. R. B. de Souza. 2016. Uma Revisão da Literatura sobre Ecossistemas de Startups de Tecnologia. In *XII Brazilian Symposium on Information Systems (Florianópolis)*. SBC, 385–392.
- [10] O. Dedeheyir, S. J. Makinen, and J. R. Ortt. 2016. Roles during innovation ecosystem genesis: A literature review. *Technological Forecasting & Social Change* (2016). <https://doi.org/10.1016/j.techfore.2016.11.028>
- [11] L. A. Digiampietri, W. T. Maruyama, C. R. N. Santiago, and J. J. S. Lima. 2015. A Link Prediction System in Social Networks. In *XI Brazilian Symposium on Information Systems (Goiania, GO)*. ACM, 139–146.
- [12] R. E. L. Escalfoni, M. A. S. Irineu, and J. Oliveira. 2017. Impacto das Redes de Negócios para Startups: Um Estudo Empírico na IETEC/CEFET-RJ. In *Encontro de Inovação em Sistemas de Informação - EISI, SBSI (Lavras)*. SBC.
- [13] T. França, F. F. Faria, C. Miceli, F. Rangel, and J. Oliveira. 2014. Big Social Data: Princípios sobre Coleta, Tratamento e Análise de Dados Sociais. In *Anais do SBBB (Porto Alegre)*. SBC, 1–40.
- [14] M. T. Hansen, H. W. Chesbrough, N. Nohria, and D. N. Sull. 2000. Networked Incubators: hothouses of the new economy. *Harvard Business Review* 78, 5 (September 2000), 74–84.
- [15] R. High. 2012. The era of cognitive systems: An inside look at IBM Watson and how it works.
- [16] R. K. Ikenami. 2016. *A abordagem ecossistema em teoria organizacional: fundamentos e contribuições*. Master's thesis. Escola Politécnica, Universidade de São Paulo.
- [17] N. Islam, P. Buxmann, and D. Ding. 2017. Fostering Digital Innovation through Inter-organizational Collaboration between Incumbent Firms and Start-ups. In *Proceedings of the 25th European Conference on Information Systems (ECIS)* (Gumarães). 1029–1043.
- [18] A. M. Magdaleno, P. Engiel, R. Tavares, P. S. Pisa, and R. M. Araújo. 2017. Bridging the Gap between Brazilian Startups and Business Processes - Process Thinkings Initial Exploratory Case Study. *iSys - Revista Brasileira de Sistemas de Informação* 10, 1 (2017), 19–38.
- [19] S. Mian, W. Lamine, and A. Fayolle. 2016. Technology Business Incubation: An overview of the state of knowledge. *Technovation* (2016). <https://doi.org/10.1016/j.technovation.2016.02.005i>
- [20] J. F. Moore. 1993. Predators and prey: A new ecology of competition. *Harvard Business Review* 71, 3 (1993), 75–83.
- [21] D. Mota, T. Nodehi, A. Zutshi, A. Grilo, and V. Cruz-Machado. 2016. What is a Digital Entrepreneurship Ecosystem: Players and Processes. In *International Conclave On Innovations in Engineering & Management (ICIEM-2016)*. 298–307.
- [22] X. Neumeyer and S. C. Santos. 2017. Sustainable Business Models, Venture Typologies, and Entrepreneurial Ecosystems: A Social Network Perspective. *Journal of Cleaner Production* (2017). <https://doi.org/10.1016/j.jclepro.2017.08.216>
- [23] B. Neves and A. A. Ferreira. 2016. An Unsupervised Topic-based Method for Identifying Reputation Dimensions in Microblogs. In *XII Brazilian Symposium on Information Systems (Florianópolis, SC)*. ACM, 33–40.
- [24] P. G. M. Oliveira and M. M. V. de Paula. 2017. Analysis of the Collaboration Network and Scope of the Brazilian Symposium on Information Systems. In *XIII Brazilian Symposium on Information Systems (Lavras, MG)*. ACM, 64–71.
- [25] L. M. Romão, M. L. Borba, and V. O. Collere. 2016. Desenvolvendo Empreendedores e Criando Startups no curso de Sistemas de Informação. *iSys - Revista Brasileira de Sistemas de Informação* 9, 3 (2016), 56–68.
- [26] J. L. Scillitoe and A. K. Chakrabarti. 2010. The Role of Incubator Interactions in Assisting New Ventures. *Technovation* 30, 3 (March 2010), 155–167.
- [27] D. P. Soetanto and S. L. Jack. 2013. Business incubators and the networks of technology-based firms. *Journal of Technology Transfer* 38, 4 (November 2013), 432–453.
- [28] B. Spigel. 2015. The Relational Organization of Entrepreneurial Ecosystems. *ET&P* 41, 1 (January 2015), 49–72.
- [29] I. Theodoro, F. Rangel, T. França, and J. Oliveira. 2014. Análise dos Influenciadores dos Protestos Brasileiros de 2013 via Twitter. In *III Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)* (Brasília). SBC.
- [30] C. Wohlin, P. Runeson, and M. Host. 2012. *Experimentation in Software Engineering*. Springer, Boston, USA.
- [31] J. Zhang, M. S. Arckerman, and L. Adamic. 2007. Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th international conference on World Wide Web*. ACM, 221–230.