Information Processing and Management 53 (2017) 1287–1298

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Robustness and stability of enterprise intranet social networks: The impact of moderators

Andrea Fronzetti Colladon*, Fabrizio Vagaggini

Department of Enterprise Engineering, University of Rome Tor Vergata, Via del Politecnico, 1, 00133 Rome, Italy

ARTICLE INFO

Article history:
Received 27 April 2017
Revised 6 July 2017
Accepted 7 July 2017

Keywords:
Enterprise intranet
Online forum networks
Moderators
Robustness
Stability
Network analysis

ABSTRACT

In this study, we tested the robustness of three communication networks extracted from the online forums included in the intranet platforms of three large companies. For each company we analyzed the communication among employees both in terms of network structure and content (language used). Over a period of eight months, we analyzed more than 52,000 messages posted by approximately 12,000 employees. Specifically, we tested the network robustness and the stability of a set of structural and semantic metrics, while applying several different node removal strategies. We removed the forum moderators, the spammers, the overly connected nodes and the nodes lying at the network periphery, also testing different combinations of these selections. Results indicate that removing spammers and very peripheral nodes can be a relatively low impact strategy in this context; accordingly, it could be used to clean the "noise" generated by these types of social actor and to reduce the computation complexity of the analysis. On the other hand, the removal of moderators seems to have a significant impact on the network connectivity and the shared content. The most affected variables are closeness centrality and contribution index. We also found that the removal of overly connected nodes can significantly change the network structure. Lastly, we compared the behavior of moderators with the other users, finding distinctive characteristics by which moderators can be identified when their list is unknown. Our findings can help online community managers to understand the role of moderators within intranet forums and can be useful for social network analysts who are interested in evaluating the effects of graph simplification techniques.

1. Introduction

Many complex systems can be described as networks in which the constituent components are represented by vertices and their connections represent the relationship existing between them. A fundamental issue concerning complex networks is the robustness of the overall system to the failure of its constituent parts. Network robustness is defined as the ability to retain one or more specific properties under perturbation of its structure, e.g. after node or edge removal (Albert, Jeong, & Barabasi, 2000; Barabasi, 2016). Robust networks are more tolerant to random failures and less vulnerable to intentional attacks against nodes and edges (Albert et al., 2000; Barrat, Bathelemy, & Vespignani, 2008). The concept of robustness is often linked to the study of network resilience, which is partially different as it usually implies the additional analysis of the network evolution over time, after potentially harmful events. Network resilience is defined as "the ability of a system to adapt to errors or intentional attacks, by changing its mode of operation without losing its ability to function" (Barabasi,

* Corresponding author.
E-mail addresses: fronzetti.colladon@dii.uniroma2.it (A. Fronzetti Colladon), fabrizio.vagaggini@gmail.com (F. Vagaggini).

http://dx.doi.org/10.1016/j.ipm.2017.07.001
0306-4573/

2016, p. 303; Gao, Barzel, & Barabási, 2016). An understanding of vulnerabilities of complex networks is fundamental to the design of robust systems (Estrada, 2006), as well as to the definition of strategies for the management of their social impact (Holme, 2004).

Following this approach, we investigated the concepts of network robustness and stability in the context of enterprise forum networks, used by employees to communicate and share knowledge. We studied the interactions taking place among the employees of three large companies, using the forums provided on the company intranets. Analyzing the interaction patterns in these three intranets, over an eight-month period, we were able to map three different networks, where employees accounts (network nodes) are linked together if they interacted in the same forum post, for instance answering to each other's comments.

1.1. Research objective

Considering the three enterprise intranets in our case study, we specifically analyzed the network robustness and the stability of well-known structural metrics, while simultaneously removing nodes based on their category (e.g. moderators or spammers), or on their structural properties, such as their degree centrality. In addition, we investigated the changes produced in the shared content, in terms of language sentiment, emotionality and complexity. Specifically, we focused our attention on forum moderators – who managed the forum content moderating discussions and posting messages which conveyed the institutional view of the company management – and on some removal strategies aimed at reducing the "noise" generated by other specific categories of nodes: the spammers, identified as those social actors who post mostly undesired and irrelevant messages; the loosely connected nodes, lying at the network periphery; the overly connected nodes, as possible alternative identification of spammers.

Our research has two main objectives: first, we want to help network analysts understand which users can be removed from intranet forum networks without significantly harming the general connectivity and the sentiment of the language used, i.e. keeping structural and semantic variables stable. Simplifying the network graph, before carrying a social network analysis, is indeed often useful to reduce the computational complexity of some algorithms and to clean the signal from possible disturbances introduced, for instance, by irrelevant messages posted by spammers. Secondly, we try to help community managers to better understand the role of moderators and their impact on the internal communication processes. Understanding the distinctive behaviors of moderators and how they influence the network connectivity can be very useful for managers who want to foster and optimize the social interaction among employees.

2. Literature review

The properties of robustness and resilience of complex networks to attacks or failures, on nodes or edges, has been widely investigated over many disciplines (Gao, Liu, Li, & Havlin, 2015), such as the World Wide Web (Albert et al., 2000; Boldi, Rosa, & Vigna, 2013; Broder et al., 2000), social and collaborative networks (Baek, Meroni, & Manzini, 2015), enterprise communication systems and e-mail networks (Aedo, Díaz, Carroll, Convertino, & Rosson, 2010; Wang, Gao, & Ip, 2010), and other disciplines like supply chain management (Tang, Jing, He, & Stanley, 2016), socio-ecological systems (Crespo, Suire, & Vicente, 2014), or biological networks (Aerts, Fias, Caeyenberghs, & Marinazzo, 2016).

There are several different ways by which vertices and edges can be removed and networks can show varying degrees of robustness to these strategies (Braunstein, Dall'Asta, Semerjian, & Zdeborová, 2016). We distinguish between network failures, which are typically attacks without a prior knowledge of the network structure or due to unexpected system errors, and targeted attacks which are usually meant to maximize the damage to the network connectivity and functioning. Nodes and edges can be removed simultaneously or sequentially (Holme, Kim, Yoon, & Han, 2002; Ventresca & Aleman, 2015). In a sequential attack, nodes or edges are progressively removed, each time observing the effect on the overall system, before deciding the next target. By contrast, a simultaneous attack is implemented at the same time against a previously selected subset of nodes or edges.

Investigations about robustness and resilience of real world networks have been centered on the analysis of several parameters such as connectivity, network distances, stability of centrality measures, average path lengths and clustering coefficients (Cohen & Havlin, 2010; Larhlimi, Blachon, Selbig, & Nikoloski, 2011). If an attacker desires to severely harm the functionality of a network, an intuitive strategy would be to target more critical social actors. Without prior information about the phenomenon represented by the network, only its structure can be examined, so the best strategies will target important nodes, such as those with a high degree or betweenness centrality (Borgatti, Everett, & Johnson, 2013).

Albert et al. (2000) sparked considerable interest in network robustness. They analyzed the impact of random failures and targeted attacks on two real-world graphs representing the topology of the Internet and a subset of the World Wide Web pages; both these graphs showed a power-law degree distribution (Broder et al., 2000). Studying the trend of the mean distance between vertices as a function of the number of removed nodes, they concluded that scale-free networks are extremely tolerant to random failures, but severely affected by targeted attacks. Broder et al. (2000) came independently to a similar conclusion studying a larger subset of the World Wide Web. Authors found that all nodes with a degree greater than five should be removed to significantly harm the graph connectivity. Accordingly, they maintained that the network was robust against targeted attacks, which seems to contrast the work of Albert et al. (2000). However, even if removing nodes with degrees greater than five seems to be a massive attack on the graph, there is no conflict between these results