Applied Mathematics and Computation 339 (2018) 651–662

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

An empirical investigation of network polarization

Ruben Interian*, Celso C. Ribeiro

Institute of Computing, Universidade Federal Fluminense, Niterói, RJ 24210-346, Brazil

ARTICLE INFO

MSC:
05C78
62F03
91F99

Keywords:
Polarization
Homophily
Network
Graph
p-Value

ABSTRACT

This paper proposes and explores a new quantitative characterization of the polarization phenomenon in networks. New tools for evaluating the polarization of a network are presented. We first characterize the homophily of each node individually. We depart from the definition of a new measure of the homophily of the nodes of a network and we consider the homophily distribution over the nodes as a primary indicator of the strength of polarization. Next, to address the polarization of the network as a whole, a probabilistic approach is developed. The approach is based on the straightforward computation of empirical cumulative distribution functions of sampled data from the network. These empirical distributions provide a more insightful understanding of the status of the network. They may be used not only to compare the polarization of groups of nodes or entire networks, but also to estimate the impacts of external interventions in terms of node polarization. The usefulness of the approach is illustrated on several case studies associated with real-life data sets from different sources.

1. Introduction

Polarization is a widely known phenomenon that has been discussed by politicians, media, and researchers in recent years [1,2]. This subject has also attracted the attention of thinkers throughout history.

Since the 19th century, Mill, an important philosopher and political theorist, claimed that dialogue across lines of political difference is a key prerequisite for sustaining a democratic citizenry [3]. Arendt also asseverated that debate is irreplaceable for forming enlightened opinions that reach beyond the limits of one's own subjectivity to incorporate the standpoints of others [4]. More recently, several world leaders have often expressed concern about polarization problems caused by social media [5,6]. From sociologists to economists, many are interested in studying the behavior and interactions in social networks that rule the opinion formation process.

According to the Oxford Dictionaries, polarization is the *division into sharply contrasting groups or sets of opinions or beliefs* [7].

It is not a widely accepted fact that social media increase polarization. Although there are many arguments supporting this thesis [8,9], there are also opposing views [10]. However, it is known that social networks and mass media, like newspapers and blogs, are the place where this phenomenon manifests itself in a more strong way. Even if social and mass media do not contribute to increase polarization in modern societies, it is important to identify the mechanisms by which polarization arises, as well as the characteristics and the peculiarities of polarized networks.

* Corresponding author.
E-mail addresses: rinterian@ic.uff.br, rubenus@yandex.ru (R. Interian), celso@ic.uff.br (C.C. Ribeiro).

https://doi.org/10.1016/j.amc.2018.07.066
0096-3003/

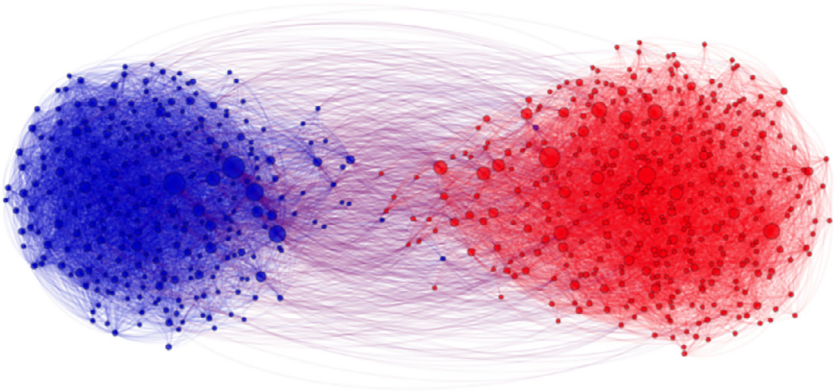

Fig. 1. Network of political blogs during the 2004 U.S. presidential election. Democratic and republican blogs are represented by blue and red circles, respectively. It is clear that this network is strongly polarized, since there are a huge number of edges connecting democratic blogs among themselves, a huge number of edges connecting republican blogs as well, and relatively few edges connecting a democratic blog and a republican blog. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Polarization is closely related to homophily (from Ancient Greek: *homo* = "self" and *philia* = "love", love to oneself), the tendency of individuals to associate with others that are similar to themselves. To avoid misunderstandings, we regard polarization as an extreme expression of homophily.

Homophily, also called assortativity [11,12], has been widely studied by researchers. Several measures and models exist to characterize homophily. However, most of the work in the literature make use of a single numerical measure to define the homophily of some network in a given moment of time. An assortativity coefficient is proposed in [12] and applied to several networks, showing that homophily is a universal phenomenon. The probability of creating a new link between two individuals as a function of their similarity is studied in [13]. The inbreeding homophily measure that reflects the amount of bias towards same-type relationships is mentioned in [14]. The Pearson correlation coefficient is used in [11] as a measure of the preference of high-degree nodes to attach to other high-degree nodes.

The characterization and the properties of polarized networks are very helpful to provide a better understanding of the behavior of individuals and societies. In addition, they can also help policymakers to define policies based on these characteristics. Although quantitative methods are often difficult to use in real-life situations because they tackle models that are abstractions of concrete cases, they are very useful in moderation systems used to detect suspicious events or users, as well as for economists and marketing experts when dealing with social and complex networks [15].

In this work, we develop tools that can be used for evaluating the polarization of a network in a more deep way than that offered by single valued measures. We first characterize the homophily of each node individually and, then, the polarization of the network as a whole. Section 2 introduces some real-life test instances that will be used as case studies. The homophily of each node is defined in Section 3 and we consider the distribution of the homophily values over the nodes of the network as a primary indicator of the strength of polarization, which is analyzed and illustrated with some case studies. The probability of a node to be influenced by homophily is derived in Section 4 as an improved measure of network polarization. This new measure is used to assess the statistical relevance of the homophily value. Section 5 develops a probabilistic approach to compare the polarization of groups of nodes or entire networks. The approach is based on the straightforward computation of empirical cumulative distribution functions of sampled data from the network, which provide a more insightful understanding of the status of the network. The usefulness of the approach is illustrated on several case studies associated with real-life data sets. Concluding remarks are drawn in the last section.

2. Case studies

We consider as case studies a number of test data sets selected from very different sources. We assume that some of them may be polarized to some extent.

- *Books* – A network of books about U.S. politics sold by Amazon.com [16]. Edges between books represent frequent copurchasing of those books by the same buyers. Most of the books are classified as conservative or liberal, and a small number of them as neutral.
- *Blogs* – A network of political blogs that emerged during the 2004 U.S. presidential election [17]. Blogs are divided into two groups: republican and democratic. Fig. 1 represents democratic and republican blogs by blue and red circles, respectively [18].
- *Trade* – A world trade network, built using information about bilateral trade data between countries. The data compiled by Aller et al. [19] was obtained from the United Nations COMTRADE database. We consider that two countries are connected by an edge if the amount of trade between them is at least 5% of the total traded by any two countries. In other words, we discard the edges that do not represent a significant trade for either country. We analyzed two groups