Computers in Human Behavior 72 (2017) 208–221

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Full length article

The effect of preference similarity on the formation of clusters and the connectivity of social networks

Madjid Tavana [a, b, *], Francisco J. Santos-Arteaga [c, d], Debora Di Caprio [e, d]

[a] Business Systems and Analytics Department, Distinguished Chair of Business Systems and Analytics, La Salle University, Philadelphia, PA, 19141, United States
[b] Business Information Systems Department, Faculty of Business Administration and Economics, University of Paderborn, D-33098, Paderborn, Germany
[c] School of Economics and Management, Free University of Bolzano, Piazza Università 1, 39100, Bolzano, Italy
[d] Instituto Complutense de Estudios Internacionales, Universidad Complutense de Madrid, Finca Mas Ferré, Edificio A, Campus de Somosaguas, 28223, Pozuelo de Alarcón, Madrid, Spain
[e] Department of Mathematics and Statistics, York University, Toronto, M3J 1P3, Canada

ARTICLE INFO

Article history:
Received 19 May 2016
Received in revised form
17 February 2017
Accepted 20 February 2017

Keywords:
Social media
Preference similarity
Expected utility
Self-organizing map
Neural networks

ABSTRACT

The current paper analyzes the formation of social networks determined by the preferences of their users, who are endowed with incomplete information regarding the characteristics of other users from who they receive friendship requests. The acceptance or rejection decision is determined by the limited information available when receiving the requests, the expectations of the users regarding the remaining characteristics of the requesters and the resulting improvement in network capacity derived from accepting the friendship requests. We illustrate how the similarity in preferences among users leads to more concentrated clusters within the incomplete information scenario analyzed. At the same time, the emergence of disutility costs derived from a suboptimal decision when accepting to interact with other users increments the dispersion between clusters. In this regard, the inclusion of requesters endowed with average preferences relative to those of the standard users composing the network acts as a connectivity-enhancing mechanism designed to reduce the dispersion and differences existing between clusters.

1. Motivation and contribution

Social media provide a substantial amount of information regarding the set of potential friends with whom one may connect after joining as a user (Adamic & Adar, 2003; Zuo, Blackburn, Kourtellis, Skvoretz, & Iamnitchi, 2016). The strategic use of the information available in social media by other users and, in particular, by companies has been consistently analyzed in the literature on social networks (Hofstra, Corten, & Buskens, 2015; Stefanone, Hurley, Egnoto, & Covert, 2015). At the same time, this literature acknowledges the existing diversity of users determined by their networking capacities and their ability to influence other users (Guo, Pathak, & Cheng, 2015; Klein, Ahlf, & Sharma, 2015).

Contrary to the standard literature on social networks, which generally focuses on analyzing the main properties of networks that have already been built (Han, Wang, Crespi, Park, & Cuevas, 2015; Jackson, 2010), we concentrate on the formation of networks and clusters within them based on the characteristics of their users. That is, consider a social medium whose users are endowed with incomplete information regarding the characteristics of other users from who they may receive friendship requests. As a result, a decision maker (DM) has to decide whether to accept a given friendship request, generating a link that may allow him to expand his current network of connections further, or reject it, expecting to find a requester who aligns better with his preferences.

If the request is accepted, then additional information becomes available regarding the characteristics of the requester. However, the acceptance decision must be made while constrained by the initial amount of incomplete information. At the same time, this information must be used by a DM to define his expectations about

* Corresponding author. Business Systems and Analytics Department, Distinguished Chair of Business Systems and Analytics, La Salle University, Philadelphia, PA, 19141, United States.
E-mail addresses: tavana@lasalle.edu (M. Tavana), fsantosarteaga@unibz.it, fransant@ucm.es (F.J. Santos-Arteaga), dicaper@mathstat.yorku.ca (D. Di Caprio).
URL: http://tavana.us/

http://dx.doi.org/10.1016/j.chb.2017.02.049
0747-5632/

the remaining characteristics of the requesters and the potential network improvements that may be achieved by accepting the request. Therefore, we will assume that the information initially available to the DM conditions the expected realizations of the remaining characteristics of the requesters, including their capacity to expand the network of connections of the DM. Consequently, our model will be designed following a decision theoretical approach based on the expected utility that could be achieved by the DMs composing a given social medium (Kahneman & Tversky, 2000; Tavana, Di Caprio & Santos-Arteaga, 2016; Tavana, Di Caprio, Santos-Arteaga, and Tierney, 2016).

The main objective of the current paper is to analyze the type of clustered structures generated within a (social) network by

- the preferences of the DMs composing the network;
- the subjective beliefs of the DMs regarding the networking capacity of the requesters;
- the disutility costs derived by the DMs from accepting the friendship of a requester whose tastes and characteristics differ significantly from their own.

We formalize the problem faced by a DM, define the expected utility tradeoffs that he faces when receiving a friendship request and simulate the resulting acceptance and rejection incentives numerically. These incentives determine the social behavior of the DM together with the structure of the resulting networks, which is based on the preferences of the DMs and their expectations regarding those of the requesters.

We build the corresponding social networks through self-organizing maps that cluster the DMs by their friendship acceptance behavior. This behavior is, at the same time, determined by the distribution of characteristics of the requesters relative to the preferences of the DMs. We illustrate how the similarity in preferences among users leads to more concentrated clusters within the incomplete information scenario analyzed. Moreover, the emergence of disutility costs derived from a suboptimal decision when accepting to interact with other users increments the dispersion between clusters. In this regard, the inclusion of requesters endowed with average preferences relative to those of the standard users composing the network acts as a connectivity-enhancing mechanism designed to reduce the dispersion and differences existing between clusters.

We should emphasize that, even though self-organizing maps do not quantify the connectivity of the resulting graphs, a visual examination will suffice to analyze their main clustering properties. In the current setting, we are interested in the concentration arising within clusters and the separation between them, both of which can be inferred from a visual examination of the weight planes and U-matrices provided by the self-organizing map algorithm.

The paper proceeds as follows. Sections 2 and 3 describe the basic structure on which the accept and reject functions are built. These functions are introduced in Section 4 and Section 5, respectively, and simulated numerically in Section 6. Section 7 analyzes the different clustered structures that arise after implementing a self-organizing map algorithm to classify the number of friendship requests accepted by the DMs. Section 8 describes the consequences from increasing the amount of information available to the DM before accepting a request. Section 9 presents some concluding remarks.

2. Basic assumptions

The choice made by the DM regarding the friendship request depends on the following variables, whose domains are also provided in the respective definitions:

- $X_1 = [x_1^m, x_1^M]$: represents the characteristics/preferences of the requester that are directly observable when receiving a friendship request. This variable accounts for the publicly available information describing the main preferences of the requester. That is, we assume that these preferences can be inferred from the interests (i.e. likes) displayed in the profile of a requester (Kosinski, Stillwell, & Graepel, 2013, 2015; Meshi, Tamir, & Heekeren, 2015). In this regard, the X_1 variable provides only part of the information required to fully infer the preferences of the requester. We will assume that the value of the realization of X_1 observed is related to the remaining information completing the profile of the requester. However, this information only becomes available after accepting the request, together with the list of friends and, therefore, the networking capacity of the requester.
- $X_2 = [x_2^m, x_2^M]$: accounts for the characteristics/preferences of the requester that become observable after accepting the friendship request. This variable provides additional information to the DM regarding the interests and tastes of the requester together with his potential networking capacity. Thus, the distribution defining the expected realization of this variable is conditioned by the observed realization of X_1. At the same time, both X_1 and X_2 will be used by the DM to infer the potential capacity of the requester to extend his network with friends whose preferences are similar to his own (Ding, Yan, Zhang, Dai, & Dong, 2016; Hu & Yang, 2015; Mislove, Viswanath, Gummadi, & Druschel, 2010).
- $X_3 = [0, 1]$: reflects the networking capacity of the requester. The shape of its associated probability function is subjectively determined by the DM based on the observed value of X_1 and the expected realization of X_2. It should be remarked that the requester can generally classify his friends in several categories, granting them access to different amounts of information depending on their degree of friendship with the requester. However, even if the DM is not granted access to the whole network, he can still benefit from the fact that other friends of the requester are actually able to observe him. That is, the DM can expand his network with friends of the requester even if they are classified in different categories.

The decision taken by the DM will be based on two incentive functions that define the expected utility derived from either accepting a given friendship request or rejecting it. Note that both these functions must be defined for the values of all the realizations of X_1 that may be initially observed by the DM. We describe these functions in detail through the following sections.

3. Utility and probability density functions

Throughout the paper, we will assume the DM to be endowed with the following utility functions and probability density functions.

- Utility function on X_i, $i = 1, 2$:

$$u_i(x_i) = x_i.$$

Hence, the first two characteristics are additively separable.

- Utility function on $X_1 \times X_2 \times X_3$:

$$u(x_1, x_2, x_3) = (x_1 + x_2)x_3.$$

The function u defines the DM's utility derived from accepting a requester with first and second characteristics given by x_1 and x_2, and networking capacity x_3. This utility plays a crucial role when constructing a decision function that allows the DM to evaluate his