Social Networks 31 (2009) 262–270

Contents lists available at ScienceDirect

Social Networks

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

Explaining the power-law degree distribution in a social commerce network

Andrew T. Stephen[a,*], Olivier Toubia[b,1]

[a] Marketing, INSEAD, Boulevard de Constance, Fontainebleau 77305, France
[b] Business, Columbia University, Graduate School of Business, 3022 Broadway, Uris 522, New York, NY 10027, USA

ARTICLE INFO

Keywords:
Network evolution
Preferential attachment
Clustering
Reciprocity
Empirical modeling
Social commerce

ABSTRACT

Social commerce is an emerging trend in which online shops create referral hyperlinks to other shops in the same online marketplace. We study the evolution of a social commerce network in a large online marketplace. Our dataset starts before the birth of the network (at which points shops were not linked to each other) and includes the birth of the network. The network under study exhibits a typical power-law degree distribution. We empirically compare a set of edge formation mechanisms (including preferential attachment and triadic closure) that may explain the emergence of this property. Our results suggest that the evolution of the network and the emergence of its power-law degree distribution are better explained by a network evolution mechanism that relies on vertex attributes that are not based on the structure of the network. Specifically, our analysis suggests that the power-law degree distribution emerges because shops prefer to connect to shops with more diverse assortments, and assortment diversity follows a power-law distribution. Shops with more diverse assortments are more attractive to link to because they are more likely to bring traffic from consumers browsing the WWW. Therefore, our results also imply that social commerce networks should not be studied in isolation, but rather in the context of the broader network in which they are embedded (the WWW).

1. Introduction

Social commerce is an emerging and fast-growing trend in which online shops (typically operated by individuals) are able to connect with other shops in the same online marketplace (e.g., Shopit.com, Squidoo.com, Zlio.com, EBay neighborhoods). Participants in these online marketplaces typically earn commissions on the sales made through their shops. These marketplaces usually comprise a large number of small shops, which enables better serving the so-called "long tail" of consumer demand (Anderson, 2006). A downside, however, is that the proliferation of shops makes it harder for any single shop to be found by consumers. Thus, accessibility becomes a key issue. Networks linking shops in such marketplaces are one way to make these shops more accessible to consumers (Stephen and Toubia, forthcoming). Links between shops in these so-called social commerce networks are usually directed, clickable hyperlinks that customers can use to move between shops. Of course, shops may also be accessed through search engines (e.g., Google), directories, and other web sites that are not part of the same marketplace.

We study the formation and evolution of a large social commerce network. This network, like many other real-world networks across various domains, has a power-law degree distribution. Our main objective in this paper is to understand how this property arose.

While a number of specific edge formation mechanisms have been discussed in the literature as giving rise to networks with power-law degree distributions (e.g., preferential attachment, triadic closure), empirical comparisons of such mechanisms have been less common. We develop a tractable (given our large dataset) dynamic statistical model that allows us to quantify the influence of various network evolution (edge formation) mechanisms that may explain the emergence of the network's power-law degree distribution. We find that the evolution of this network and the emergence of its power-law degree distribution are consistent with a mechanism that relies on vertex attributes that are not based on the structure of the network: sellers prefer to connect their shops to shops that have more diverse product assortments, where assortment diversity follows a power-law distribution (over shops). Shops with more diverse assortments are more attractive to link to because they are more likely to bring traffic from consumers browsing the WWW. Therefore, our results also imply that social commerce networks should not be studied in isolation, but rather in the context of the broader network in which they are embedded (the WWW).

This paper is organized as follows. We describe the context of social commerce and our specific dataset in Section 2. In Section 3 we show that the network has a power-law degree distribution. In Section 4 we discuss and formalize a set of social network evolution mechanisms that may explain the emergence of the network's

* Corresponding author. Tel.: +33 01 60 71 26 47; fax: +33 01 60 74 55 00.
E-mail addresses: Andrew.Stephen@insead.edu (A.T. Stephen), ot2107@columbia.edu (O. Toubia).
[1] Tel.: +1 212 854 8243; fax: +1 212 854 7647.

0378-8733/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.socnet.2009.07.002

power-law degree distribution. We quantify the impact of these mechanisms on the evolution of the social network in Section 5, and conclude in Section 6.

2. Data and context

Our data come from an online marketplace that leverages the "affiliate" programs offered by online retailers (i.e., commissions offered to web sites that refer customers to the retailers). Members of the marketplace ("sellers") create personalized shops (each shop has its own URL) and add products to their shops from a database of over 4 million products. These products are actually sold by over 100 partner online retailers (who act as vendors/wholesalers). Sellers may add or remove products from their shops at any time. Sellers are individual people, not companies, and they do not hold inventory, set prices, or advertise directly. Instead, when a customer purchases a product from a shop, he or she is directed to a checkout page on the vendor's website. The vendor processes the transaction and pays a commission to the company that hosts the marketplace. The company in turn redistributes part of this commission to the seller whose shop generated the sale.

The way product assortments are organized in this marketplace will be important in our analysis. Each time a seller adds a new product to his or her shop, he or she has to assign this product to a category. Possible categories include a set of generic categories proposed by the company hosting the marketplace (e.g., "apparel," "books"), and seller-created categories (e.g., "My favorite books on social networks"). A given shop's front page consists of K equal-sized "boxes," one for each of the shop's K categories. The size of each box is common across all shops in the marketplace. Each box contains a short description of at most three products listed in the corresponding category, and a thumbnail picture of the first product listed in that category. If more than three products are offered in a given category, the box corresponding to that category contains a "view all products in that category" link. See Fig. 1 for an illustration. This is the standard layout for every shop, and sellers can control only the products and their categories (prices are set by vendors, not sellers; webpage layout is set by the marketplace owner). The page layout interestingly makes the number of categories a shop has very salient, even at first glance.

Approximately 18 months after this marketplace was established, the company introduced a new social feature that allowed sellers to post hyperlinks from their shops to any other shop in the marketplace. A social commerce network was thus born with shops as vertices, and hyperlinks between them as directed edges. Prior to this feature being introduced, shops were independent. Links are publicly displayed on each shop's homepage, participation in this network is not compulsory, and links are permanent (i.e., they cannot be removed, which means for example that sellers cannot threaten link-removal as a strategic device against other sellers). The links here act as referrals in the sense that a link posted on shop i pointing to shop j refers customers who visit shop i to shop j. These links are simply Internet hyperlinks and they carry no additional information other than the name of the to-shop. When shop i creates a link to shop j, shop j is made aware of this link by an email that contains a hyperlink to i, offering j an easy opportunity to reciprocate i's link. Reciprocating links is optional, therefore the edges are directed. The fact that the network was born after the marketplace was already well-established implies that we have access to a set of relevant measures that describe a given shop *before* it entered the network. These measures are completely exogenous to the network (i.e., the network had no influence on them).

Our dataset covers 2 years of activity in this marketplace from its inception, beginning 18 months before and ending 24 weeks after the birth of the network. It includes network-relevant information as well as information on shops, with the network-relevant information starting at the birth of the network and the information on shops starting 18 months before the birth of the network. Over the observation window, 72,294 links were created. The number of shops in the marketplace grew to 136,774 by the end of the observation window (the marketplace contained 74,291 shops when the network was born). In terms of the network, by the end of our 24-week observation window 19,125 shops either sent or received at least one link.[2]

Stephen and Toubia (forthcoming) show that a key effect of creating a social commerce network is making shops more accessible to customers browsing the marketplace. These links allow customers to move more easily throughout the online marketplace, like in a virtual shopping mall, and shops whose accessibility is enhanced by the network earn higher revenues. We assume that sellers' decisions are targeted toward increasing the commissions earned by their shops. Therefore, we expect sellers to create links that they believe are likely to improve the accessibility of their shops.

One apparent contradiction is that links are directed hyperlinks that direct customers from one's shop to *another* shop. Therefore the primary effect of links is to drive customers *away* from one's shop and to make *other* shops more accessible. The accessibility of a shop is increased when *other* shops create incoming links to that shop. Stephen and Toubia (forthcoming) show that although outgoing links have a negative effect on commission revenue and incoming links have a positive effect, the net effect of creating a link is positive, for the following two reasons. First, the magnitude of the positive effect of an incoming link is larger than that of the negative effect of an outgoing link. Second, a large proportion of links are reciprocated. As a result, although creating a link has an immediate negative impact, it is likely to result in a reciprocated incoming link which will have a larger positive effect, making the decision to create a link consistent with profit-maximizing objectives. This implies that reciprocity is likely to play a key role in the evolution of such a network.

3. Power-law degree distribution

One of the most documented aggregate network properties is the so-called "scale free" property, which is satisfied when vertices' numbers of edges (degrees) are power-law distributed, resulting in few vertices having many edges and many vertices having few edges (Barabási and Albert, 1999; de Solla Price, 1965). This property has been found in networks in fields as diverse as Internet servers, pages on the World Wide Web, and scientific citations. Formally, a network is said to have a power-law degree distribution when for degree k, the probability distribution of k follows a power-law, i.e., $p(k) \propto k^{-\gamma}$, where $p(\cdot)$ indicates the probability mass function, and $\gamma \geq 1$ is the parameter of the power-law distribution.

Fig. 2 shows histograms of the indegree distribution in our network at various points in time, as well log–log plots of the same distribution. The histograms show a characteristic power-law "long tail," and all log–log plots are close to linear, suggesting that our network appears to have a power-law degree distribution. One standard way to estimate the power-law parameter (γ) would be to fit a linear regression to the log–log plots. Given the shortcomings of this approach (see, for example, Jones and Handcock, 2003), more appropriate approaches have been proposed, based on maximum likelihood. In particular, we used the approach proposed by Clauset

[2] On average, connected shops feature more products (214.85 versus 53.36 for disconnected shops), have more categories (23.75 versus 15.90 for disconnected shops), make more sales (87.98 versus 5.08) and earn higher commissions (€8.61 versus €0.50).