

**THE QUEST FOR CONTENT:  
THE ROLE OF PRODUCT NETWORKS AND SOCIAL  
NETWORKS IN ONLINE CONTENT EXPLORATION**

**Jacob Goldenberg**

School of Business Administration, Hebrew University, Jerusalem,  
and Columbia Business School, Columbia University, New York

**Gal Oestreicher –Singer**

Faculty of Management, Tel Aviv University

**Shachar Reichman**

Faculty of Management, Tel Aviv University

Extended Abstract of Research in progress

Submitted to WISE 2009, Phoenix, AZ

## Introduction

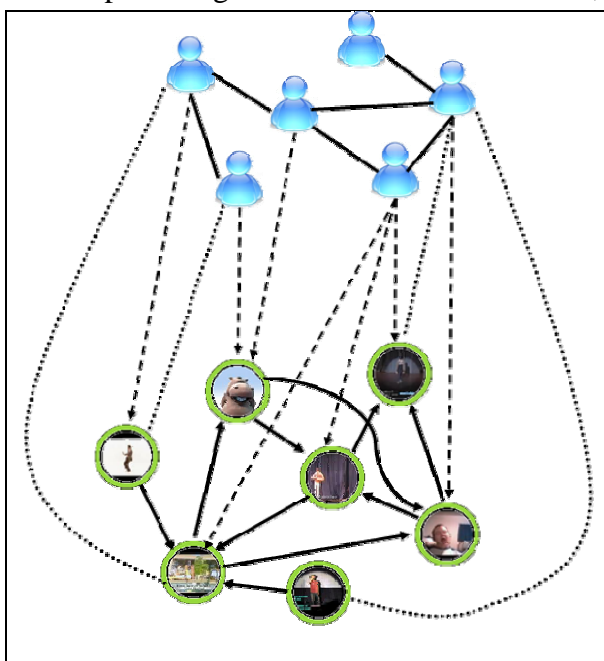
With the growth in online content consumption, the variety of offered content and the amount of content available are becoming larger than ever. Considering that online content is rarely accompanied by traditional marketing campaigns and advertising, a fundamental question that needs to be answered in this context is, how do consumers find "good" content? Additionally, from the perspective of a website that offers content, it remains to be determined which website structures facilitate efficient and successful exploration of the content space.

Recent growth in online social interaction has led to the emergence of a great number of visible electronic networks. Some represent relationships among individuals who are friends, colleagues, or trading partners. Such networks can be described as collections of personal webpages, linked by hyperlinks, and are often referred to as *social networks*. *Product networks*, in contrast, map complementarity relationships among products that are frequently consumed by overlapping sets of consumers. These networks can be described as collections of linked webpages that each represents a distinct product. Most electronic commerce sites feature such product networks. Recently, a growing number of websites have begun to offer a combination of product and social networks, linked by hyperlinks, thus creating a *dual network* structure. The dual network structure is an emerging structure in electronic commerce sites, currently most prevalent in user-generated content (UGC) sites. Our focus in this paper is on this dual network structure and its role in facilitating user exploration of content.

We focus on *ill-defined exploration* of the content space. Ill-defined exploration occurs when a consumer explores the space of options with no specific pre-defined target in mind. The exploration process continues until the consumer finds an object that matches his or her taste, or until search costs reach a specific level. It seems that a substantial portion of online content consumption is a result of ill-defined exploration. That is, consumers browse the content space without a specific target, looking for a "good" product. However, the growing amount of available online content makes the ill-defined exploration process increasingly challenging.

In this work we study the effect of the website's structure on exploration efficiency and on the consumer's overall satisfaction.

We first analyze the YouTube.com dual network, the largest available dual network. Using data on more than 700,000 videos (with 9 million links between them) and 50,000 users (with 120,000 links between them), we are able to show that the structural properties of user pages are very different from those of product pages. Specifically, the betweenness of user pages is



**Figure 1. Illustration of the YouTube.com Dual-Network**

The dotted lines represent owner links between product nodes (circles) and social network nodes (icons); dotted arrows represent favorite links; solid lines represent social links between users, and solid arrows represent "related videos" links.

three times that of product pages; thus, user pages act as brokers in the network. The second part of our work uses simulation analysis to compare and contrast the effects of dual networks, product networks, and sponsored (randomized) networks on the exploration process. We show that while random rewiring of the product network reduces the average distance between products, the bridging effect of the user pages cannot be artificially replicated. Finally, we introduce an experiment in which consumers are offered a YouTube-based website and are exposed to one of the following three treatments: recommendations based solely on the YouTube product network; recommendations based on the dual network; and recommendations based on the product network with additional random rewiring. We show that the dual network structure leads to faster access to “good” content and overall higher satisfaction. Put together, our work provides important insights regarding UGC and electronic commerce websites, suggesting that the integration of social and product networks will improve consumers' exploration of the product space and will lead to higher consumer satisfaction.

In what follows we will briefly describe our data and our salient results.

## Overview of data

Using data from YouTube.com, the largest existing dual network, we are able to conduct an in-depth analysis of the dual network structure. YouTube's core business is centered around videos, the website's “products”. Each video has an associated webpage that is connected by hyperlinks to other video webpages, thus creating a product network. In addition to the product network, YouTube offers a social network in which each user has an associated webpage; these webpages can be linked to other user pages (creating a social network) and to video pages (connecting the product and social networks). Conceptually, this creates a dual network structure, as illustrated in Figure 1.

We collected data on the YouTube.com product and social networks. Thus far we have collected data for approximately 700,000 videos and for 50,000 users connected by approximately 10 million hyperlinks. The hyperlinks are divided into three types as follows: (1) links within the product (video) network; these include directed hyperlinks appearing under the label “*Related Videos*” on a product's page and are based on co-consumption; (2) links within the social network, including undirected hyperlinks between each user and his or her friends, who are listed under “*Friends*” on the user's page; and (3) links connecting the product and social networks. These include two types of links: “*Owner*” (undirected links that link videos to their creator's page) and “*Favorites*” (directed links that link user pages to favorite videos).

## The topology of the YouTube dual network

The first step in our empirical investigation requires an in-depth analysis of the YouTube dual network structure. We computed several indices that are commonly used in the literature to characterize network structures (Wasserman and Faust, 1994; Newman, 2003), including indegree and outdegree, clustering coefficient, assortative mixing, density, distance between any two nodes in the network, Page Rank, betweenness centrality (a measure of the number of shortest paths to which the node belongs), and closeness centrality (the average shortest path between any two nodes).

We compared the summary statistics of the social and the product networks and observed that the two networks were quite different in their structures (space limitations prevent us from reporting on all the results). Subsequently, we combined the two networks into a single two-mode network with two different types of nodes: videos and users. All the different types of links mentioned above were included in this network. We recomputed the above-mentioned network indices for each node in this integrated network. We then compared the node characteristics of the two different types of nodes (i.e., video nodes and user nodes). The results are presented in Table 1.

Surprisingly, even when both types of nodes are included in the same network, it is clear that compared with product nodes, user nodes have a significantly higher betweenness despite their significantly lower outdegree. Simply put, these findings suggest that user pages play an important role in increasing network connectivity. These results are even more striking when contrasted with the results concerning closeness centrality. We see a very small difference in the average closeness centrality of the video and user nodes. Taken together, these findings imply that although the user and the video nodes are located at equal average distances from all other nodes, user nodes play an important role in bridging different parts of the network and possibly facilitating faster exploration of the content space.

Network indices	Product pages	User pages	T-test
Closeness	6.25	6.75	0.00***
Betweenness	3.89E+06	12.8E+06	0.00***
InDegree	14.25	3.97	0.00***
OutDegree	14.50	4.49	0.00***
Clustering Coefficient	0.20	0.10	0.00***

**Table 1.** The dual network indices divided by the two types of nodes

The question that should be addressed in this context is, is this bridging ability unique to user pages, or can increased connectivity be achieved simply by adding random links bridging different parts of the product network (this is often referred to as "rewiring")

**Network simulation results: Comparing different network structures**

To answer the above stated question, we compared three different network structures containing the same number of nodes: a product network, a dual network, and a product network

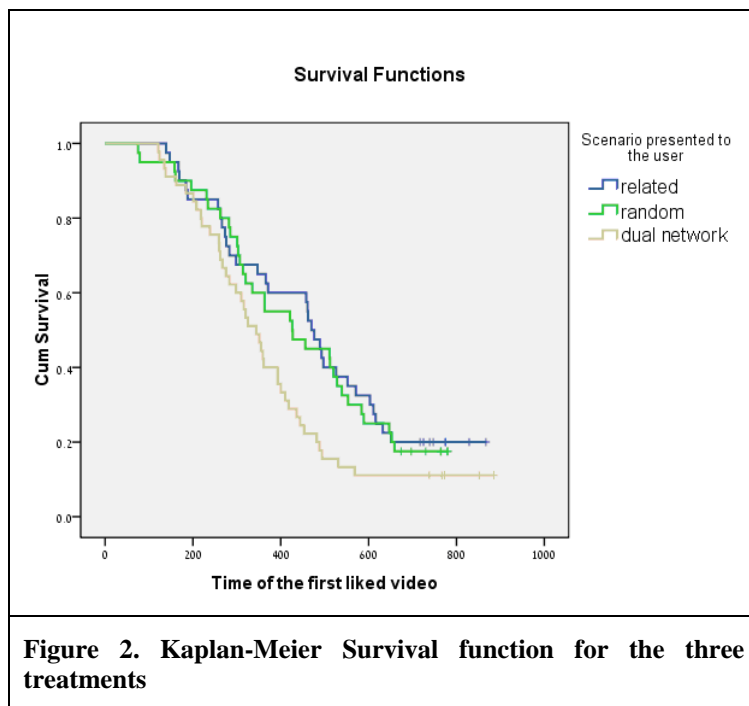
with random rewiring. In accordance with the literature, random rewiring reduced the average distance between products. However, the betweenness of the user pages in the dual network was about two times that of the added random nodes. This suggests that the self-organizing user pages encompass inexplicit characteristics that cannot be imitated by an artificial random rewiring. Users are able to create structural variety on their pages and group various videos together, reflecting the fact that different videos may be related according to a particular user's perception.

**Ill-defined exploration: An experiment**

Our empirical results show that user pages have a central position in the dual network and act as brokers between different parts of the product network. These results suggest that social nodes have an important role in facilitating ill-defined exploration. To test this conjecture, we constructed a YouTube-like website. The website provides video pages that can be viewed using a built-in video player. Each page contains ten links to other videos, enabling navigation through the site. The site offers more than 500,000 videos with about 10 million links connecting them. The videos and recommendations are taken from YouTube and are unchanged.

Participants were asked to watch videos for about 15 minutes and assign a rating from one star ("Poor") to five stars ("Awesome!") to each video they watched. Once a video had been watched, the participant was free to choose the next video out of a list of recommendations. We assigned the participants into three treatments: participants offered recommendations based on the product network only ("related videos" as suggested by YouTube); participants offered recommendations based on both the product network and the social network (a dual network); and participants offered recommendations based on the product network and randomly suggested videos (labeled "featured videos").

To study the effects of the different offered structures, we used survival analysis (and a Cox regression). The event was a successful conclusion of the exploration process (rating of "4" or "5"). The results are presented in Figure 2 and clearly show that consumers presented with a dual network find "good" content faster. We also find that their overall satisfaction with the website is higher.



**Figure 2. Kaplan-Meier Survival function for the three treatments**

These results represent the structural variety of human taste and add to our previous findings by suggesting that user pages are a valuable asset to the navigation process on the website. Individuals seem to be interested in videos from different fields and from different parts of the product network, and therefore the user pages play an important role in bridging between videos and facilitating efficient exploration of the content space.

### Concluding remarks

This work attempts to study the dual network structure and its influence on online content exploration. The presence of hyperlinked networks is one of the principal differences between the online and traditional content environments. Hyperlinked networks facilitate consumers' exploration among products, for example by diminishing the physical limitations that affect a search for products located in different stores. At the same time, as a result of the huge available variety and number of products accessible in online content environments, product searches have become much more complicated tasks. This research enhances our

understanding of how dual networks, which are grounded in the online hyperlinked environment, influence people's exploration of the content space.

Our first set of results characterizes the dual network structure on YouTube.com and demonstrates that the social network and the product network have very different characteristics. Specifically, we find that the betweenness of user pages is three times higher than that of product pages; thus, user pages bridge between different parts of the network. We also show that while random rewiring of the network reduces the average distance between products, the bridging effect of the user pages cannot be artificially replicated. Finally, we introduce an experiment in which consumers browse a YouTube-based website. We find that the dual network structure leads to a faster path to "good" content and overall higher satisfaction. Our work provides important insights regarding UGC and electronic commerce websites, suggesting that the integration of social and product networks will improve consumers' exploration of the product space and will lead to higher consumer satisfaction.

*References available on request.*