

## Chapter 2

# A Study of Homophily on Social Media

**Halil Bisgin**

*University of Arkansas at Little Rock, USA*

**Nitin Agarwal**

*University of Arkansas at Little Rock, USA*

**Xiaowei Xu**

*University of Arkansas at Little Rock, USA*

### **ABSTRACT**

*Similarity breeds connections, the principle of homophily, has been well studied in existing sociology literature. Several studies have observed this phenomenon by conducting surveys on human subjects. These studies have concluded that new ties are formed between similar individuals. This phenomenon has been used to explain several socio-psychological concepts such as segregation, community development, social mobility, etc. However, due to the nature of these studies and limitations because of involvement of human subjects, conclusions from these studies are not easily extensible in online social media. Social media, which is becoming the infinite space for interactions, has exceeded all the expectations in terms of growth, for reasons beyond human comprehension. New ties are formed in social media in the same way that they emerge in the real world. However, given the differences between real-world and online social media, do the same factors that govern the construction of new ties in the real world also govern the construction of new ties in social media? In other words, does homophily exist in social media? In this chapter, the authors study this highly significant question and propose a systematic approach by studying two online social media sites, BlogCatalog and Last.fm, and report our findings along with some interesting observations.*

DOI: 10.4018/978-1-61350-513-7.ch002

## INTRODUCTION

Originally proposed by Lazarsfeld and Merton (1954), the term *homophily* refers to the concept that similar individuals are assumed to associate with each other more often than with others. Since then hundreds of studies have been performed as summarized in (McPherson, Smith-Lovin, & Cook, 2001) that extensively investigated the phenomenon of homophily. Over the years, sociologists have studied the human population on numerous sociodemographic dimensions including race, gender, age, social class, and education and have concluded that friends, co-workers, colleagues, spouses, and other associations tend to be more similar to each other than randomly chosen members of the same population. This phenomenon has been widely used to explain certain sociological concepts like segregation, social mobility, etc.

All these studies have one thing in common, that is all of them were conducted in a physical world scenario by surveying a group of human subjects. Often these subjects belonged to a specific geographical location. These subjects were studied over a set of sociodemographic dimensions as mentioned above. Their ties were subject to social influence. For example, parents had to approve their kids' friends, individuals usually acquainted with those either in the same workplace, schools, etc. that inherently favored the conclusions of the study. Lack of a platform where individuals can explore relations outside their geographical locations, outside their social circles, outside their workplace or schools etc., made it difficult to generalize the results.

Homophily was categorized by Lazarsfeld and Merton (1954) into status homophily and value homophily. Status homophily takes into consideration the social status of individuals, implying that individuals with similar social status tend to associate with each other. Value homophily defines similarity based on what people think, implying that individuals who think alike tend to associate

with each other regardless of the differences in their social status. However, results from both the categories were concluded on a small-scale experiment involving human subjects in a physical world scenario.

Often on social media, information such as age, gender, education, and social status is either unavailable or untrustworthy. Moreover, individuals share their interests, likes, dislikes, opinions, perspectives, thoughts, etc. Due to the absence of sociodemographic dimensions, it is difficult to assume that homophily that was studied on sociodemographic dimensions. Interests of individuals are one of the strongest factors to evaluate homophily in the virtual world. These factors were often neglected in the studies conducted in the physical world. Precisely due to this reason it is difficult to evaluate status homophily in the virtual world. Authors in (Kossinets & Watts, 2009) study a university campus social network and conclude that social ties are often influenced by triadic and focal closures.

Another major difference between studies conducted in the physical and virtual worlds is the scale of the study. Millions of individuals could be easily studied in the virtual world as compared to the physical world. This makes the results much more conclusive and generalizable. Next, we summarize the differences between the physical and virtual worlds in Table 1.

Inspired by the differences between physical and online world, in this chapter we study the existence of homophily in online social networks. Specifically, we make the following contributions:

1. We study the differences between real world and online/virtual world interactions;
2. Based on the differences highlighted above, we investigate the question whether creation of new ties in the online world is inspired by homophily. In other words, we investigate whether individuals in the online world are likely to become friends if they share similar interests;

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/study-homophily-social-media/61509](http://www.igi-global.com/chapter/study-homophily-social-media/61509)

## Related Content

---

### Interestingness Measures for Association Rules: What Do They Really Measure?

Yun Sing Koh, Richard O'Keefe and Nathan Rountree (2008). *Data Mining and Knowledge Discovery Technologies* (pp. 36-58).

[www.irma-international.org/chapter/interestingness-measures-association-rules/7512](http://www.irma-international.org/chapter/interestingness-measures-association-rules/7512)

### Application of Improved Chameleon Swarm Algorithm and Improved Convolution Neural Network in Diagnosis of Skin Cancer

Wu Beibei and Nikolaj Jade (2023). *International Journal of Data Warehousing and Mining* (pp. 1-16).

[www.irma-international.org/article/application-of-improved-chameleon-swarm-algorithm-and-improved-convolution-neural-network-in-diagnosis-of-skin-cancer/325059](http://www.irma-international.org/article/application-of-improved-chameleon-swarm-algorithm-and-improved-convolution-neural-network-in-diagnosis-of-skin-cancer/325059)

### Predictive Modeling Versus Regression

Patricia Cerrito (2010). *Text Mining Techniques for Healthcare Provider Quality Determination: Methods for Rank Comparisons* (pp. 110-152).

[www.irma-international.org/chapter/predictive-modeling-versus-regression/36635](http://www.irma-international.org/chapter/predictive-modeling-versus-regression/36635)

### Efficient and Effective Aggregate Keyword Search on Relational Databases

Luping Li, Stephen Petschulat, Guanting Tang, Jian Pei and Wo-Shun Luk (2012). *International Journal of Data Warehousing and Mining* (pp. 41-81).

[www.irma-international.org/article/efficient-effective-aggregate-keyword-search/74755](http://www.irma-international.org/article/efficient-effective-aggregate-keyword-search/74755)

### ASCCN: Arbitrary Shaped Clustering Method with Compatible Nucleoids

Renxia Wan, Lixin Wang and Xiaoke Su (2010). *International Journal of Data Warehousing and Mining* (pp. 1-15).

[www.irma-international.org/article/asccn-arbitrary-shaped-clustering-method/46940](http://www.irma-international.org/article/asccn-arbitrary-shaped-clustering-method/46940)