COMMUNITY STRUCTURE IN ONLINE COLLEGIATE SOCIAL NETWORKS

AMANDA L. TRAUD[1], ERIC D. KELSIC[2], PETER J. MUCHA[1,3], AND MASON A. PORTER[4]
[1]CAROLINA CENTER FOR INTERDISCIPLINARY APPLIED MATHEMATICS, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL, NC 27599-3250, USA
[2]DEPARTMENT OF SYSTEMS BIOLOGY, HARVARD MEDICAL SCHOOL, HARVARD UNIVERSITY, BOSTON, MA 02115, USA
[3]INSTITUTE FOR ADVANCED MATERIALS, NANOSCIENCE & TECHNOLOGY, UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL, NC 27599, USA
[4]OXFORD CENTRE FOR INDUSTRIAL AND APPLIED MATHEMATICS, MATHEMATICAL INSTITUTE, UNIVERSITY OF OXFORD, OX1 3LB, UK

Abstract. *We apply the tools of network analysis to study the roles of university organizations and affiliations in structuring the social networks of students by examining the graphs of Facebook "friendships" at five American universities at a single point in time. In particular, we investigate each single-institution network's community structure, which we obtain by partitioning the graphs using an eigenvector method. We employ both graphical and quantitative tools, including pair-counting methods that we interpret through statistical analysis and permutation tests, to measure the correlations between the network communities and a set of self-identified user characteristics (residence, class year, major, and high school). We additionally investigate single-gender subsets of the university networks and also examine the impact of incomplete demographic information in the data. Our study across five universities allows one to make comparative observations about the online social lives at the different institutions, which can in turn be used to infer differences in offline lives. It also illustrates how to examine different instances of social networks constructed in similar environments, while emphasizing the array of social forces that combine to form simplified "communities" obtainable by the consideration of the friendship links. In an appendix, we review the basic properties and statistics of the employed pair-counting similarity coefficients and recall, in simplified notation, a useful analytical formula for the z-score of the Rand coefficient.*

1. Introduction. Social networks are a pervasive part of everyday life. Although they have long been studied by social scientists [101], the mainstream awareness of their ubiquity has arisen only recently, in part because of the rise of social networking sites (SNSs) on the World Wide Web. Since their introduction, SNSs such as Friendster, MySpace, Facebook, Orkut, LinkedIn, and hundreds of others have attracted millions of users, many of whom have integrated SNSs into their daily lives to communicate with friends, send e-mails, solicit opinions or votes, organize events, spread ideas, find jobs, and more [9]. As recent work has demonstrated, the implications and importance of online social networks are diverse and signicant [60], in part because of their role as a reflection of offline society [9,30,34,35,43,79]. Meanwhile, the scientic study of real-world networks, including the study of SNSs, has expanded in recent years, and there is strong optimism that formal statistical and graph-theoretic analysis can not only help achieve a better understanding of the structure and dynamics of online social networks but also inform useful additions, modications, and uses of such services.

The plethora of available activities in SNSs underscores the fact that online social networks include multiple types of interacting informational and social structures: