Are technological and social networks really different?

Daniel E. Whitney  
Engineering Systems Division  
Massachusetts Institute of Technology  
Cambridge, MA 02139  
dwhitney@mit.edu

David Alderson  
Operations Research Department  
Naval Postgraduate School  
Monterey, CA 93943  
dlalders@nps.edu

The use of the Pearson coefficient (denoted $r$) to characterize graph assortativity has been applied to networks from a variety of domains. Often, the graphs being compared are vastly different, as measured by their size (i.e., number of nodes and arcs) as well as their aggregate connectivity (i.e., degree sequence $D$). Although the hypothetical range for the Pearson coefficient is $[-1, +1]$, we show by systematically rewiring 38 example networks while preserving simplicity and connectedness that the actual lower limit may be far from $-1$ and also that when restricting attention to graphs that are connected and simple, the upper limit is often very far from $+1$. As a result, when interpreting the $r$-values of two different graphs it is important to consider not just their direct comparison but their values relative to the possible ranges for each respectively. Furthermore, network domain ("social" or "technological") is not a reliable predictor of the sign of $r$. Finally, we show that networks with observed $r < 0$ are constrained by their $D$ to have a range of possible $r$ which is mostly $< 0$, whereas networks with observed $r > 0$ suffer no such constraint. Combined, these findings say that the most minimal network structural constraint, $D$, can explain observed $r < 0$ but that network circumstances and context are necessary to explain observed $r > 0$. 
1 Introduction

Newman [1] observed that the Pearson degree correlation coefficient $r$ for some kinds of networks is consistently positive while for other kinds it is negative. Several explanations have been offered [2, 3]. In this paper we offer a different explanation based on embedding each subject network in the set of all networks sharing the subject network's degree sequence (denoted here as $D$).

Our primary contribution is to show with 38 example networks from many domains that the degree sequence for simple and connected graphs dictates in large part the values of $r$ that are possible. More precisely, we show that, although $D$ does not necessarily determine the observed value of $r$, it conclusively determines the maximum and minimum values of $r$ that each subject network could possibly have, found by rewiring it while preserving its $D$, its connectedness, and its simpleness. Approaching the problem this way reveals interesting properties of $D$ that affect the range of possible values of $r$. In particular, networks with observed $r < 0$ have a smaller range that is all or mostly $< 0$. But for networks with observed $r > 0$ the range covers most of $[-1, +1]$. After studying these properties and their underlying mathematics, we ask if the alternate wirings are semantically feasible, in an effort to see how the domain of each network might additionally constrain $r$.

2 Observed data and mathematical analysis

Table 1 lists the networks studied and their properties of interest. The values of $r_{max}$ and $r_{min}$ were obtained by systematically rewiring each subject network while preserving connectivity and degree sequence. This type of rewiring procedure was used previously by Maslov et al. [4], who argued that graph properties such as assortativity only make sense when the graph of interest is compared to its "randomized" counterpart. The message of this paper is similar in spirit, but focuses on empirical evidence across a variety of domains.

The networks in Table 1 are listed in ascending order of $r$. It should be clear from this table that one find networks of various types, such as "social," "biological," or "technological," having positive or negative values of $r$. This indicates that networks do not "naturally" have negative $r$ or that any special explanation is needed to explain why social networks have positive $r$. All empirical conclusions drawn from observations are subject to change as more observations are obtained, but this is the conclusion we draw based on our data.

In Table 1, the kinds of networks, briefly, are as follows: social networks are coauthor affiliations or clubs; mechanical assemblies comprise parts as nodes and joints between parts as edges; rail lines comprise terminals, transfer stations or rail junctions as nodes and tracks as edges; food webs comprise species as nodes and predator-prey relationships as arcs; software call graphs comprise subroutines as nodes and call-dependence relationships as arcs; Design Structure

---

1 No causality is implied. The domain may well provide the constraints that shape $D$. The present paper does not attempt to assign a causal hierarchy to the constraints.