Networks and Community Detection

Joshua Mankelow

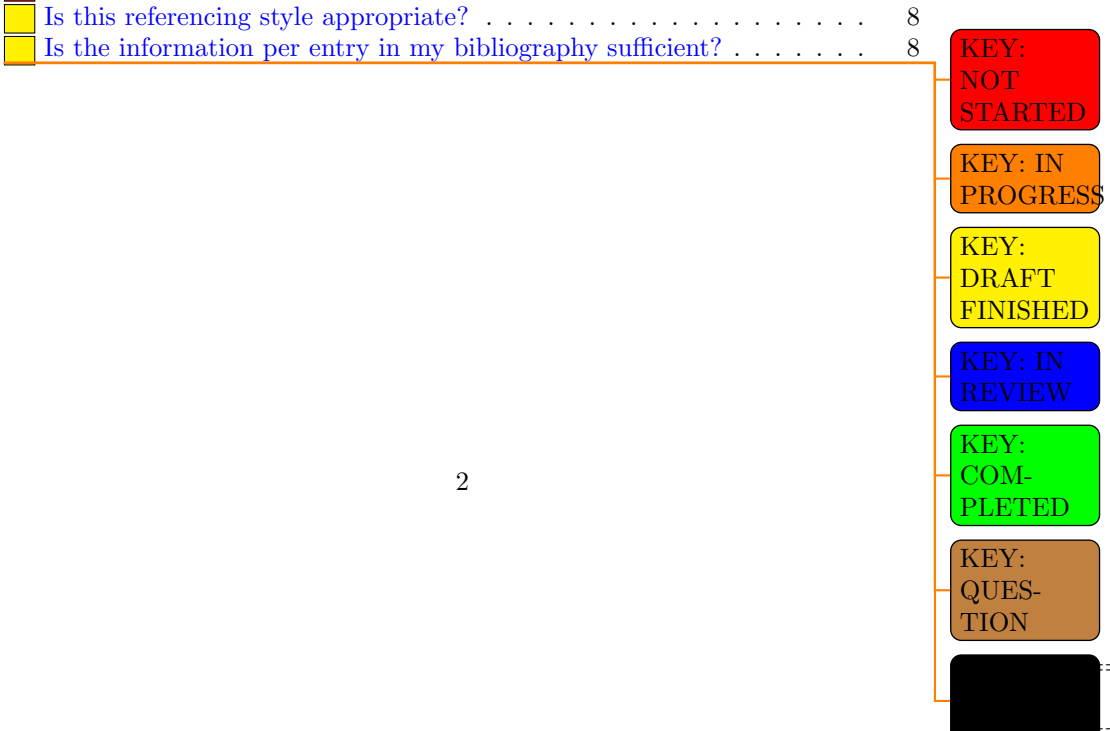
February 1, 2022

Abstract

Networks are a great model for analysing and understanding real world complex systems. A very important feature of these network models is *community detection*. Community detection allows us to identify clusters in the network that are well connected amongst eachother. If a network is being used to model a real world system then finding this structure has many implications about the behaviour of the system. This essay will discuss multiple methods for community detection in networks and their applications to the analysis and understanding of real world complex systems.

Contents

Todo list



1 Introduction to Networks

SEC: Introduce Network Theory

Networks in the technical sense are analogous to networks in the nontechnical sense - a collection of objects paired with a number of connections that can link any two objects. *Networks: An Introduction* by *Mark Newman* lists *Technological Networks*, *Social Networks*, *Information Networks* and *Biological Networks* as different systems that are modelled by the technical interpretation of a network.[New10, Contents]. A brief example of a network would be something like the following: Imagine you and your friends are represented as dots (nodes or vertices) on a piece of paper. Then if any two people are friends, the dots representing those people are connected by a line (edge). If you then repeat this process by asking your friends to list all their friends and so on, you will end up with a simple model of a *social network*.

As one might imagine, now that we have this model it's easy to be curious about any structure that emerges that we can detect and abuse to develop an understanding of the real world system that we are representing. The structure that this essay will explore is that of *communities*. Vaguely speaking, communities are subsets of a network that are *densely connected* amongst themselves. I.e. there is some notion of any node within a community being more closely connected to other nodes in the community than nodes outside the community in the average case. Before we dive into the details of communities and detecting them, I wish to provide some motivation by way of example of the kinds of situations that networks can arise and why they are the natural model for the related systems.

1.1 Social Networks

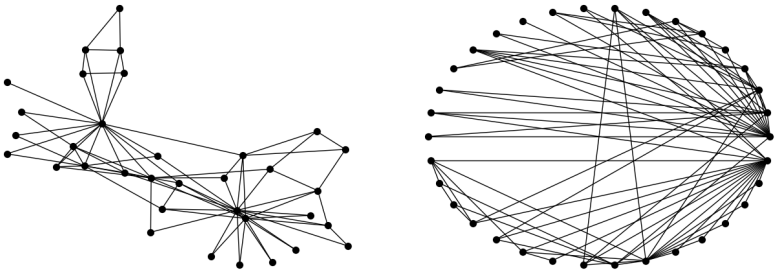
SEC: Models of Networks

To better illustrate the simple notion of a social network mentioned above, I will introduce the canonical community detection example of *Zachary's Karate Club*. Zachary's Karate Club is a dataset where "The data was collected from the members of a university karate club by Wayne Zachary in 1977. Each node represents a member of the club, and each edge represents a tie between two members of the club."[kon17, Metadata]. In Figure 1, there are two different renderings of the Zachary Karate Club. Figure 1a shows the network rendered using a "spring" layout (which is a type of force directed graph drawing[Kob12])and figure 1b shows the network rendered using a "circle" layout. These different layouts show us different parts of the underlying structure of the network. For example, in Figure 1a, it's clear which nodes in the network have the highest degree and which are of lower degree. It also allows you to see some of the community structure in the network. Meanwhile, in Figure 1b, it's much easier to see the which nodes edges in the network would need to be removed to disconnect the network in a minimal way. The reason this dataset is the canonical example of community detection is that the question that comes with it is the following: Suppose two members of the club have a disagreement which causes the club to split in two. How does the club split? In Zachary's original paper on the topic *An Information Flow Model for Conflict in Small Groups*[Zac77] he uses community detection techniques to predict how the network will split after the disagreement. Out of 34 people, Zachary correctly predicts how 33 of them will choose a side after the disagreement.

Am I supposed/allowed to reference like this?

Should I include any code that I write too even if it's just a small example like this?

There, of course, exist different ways to represent social networks. The way



(a) Spring Layout (b) Circle Layout

Figure 1: Two renderings of the Zachary Karate Club network using data from KONECT.cc[Kun13] and a Python library NetworkX[HSS08]

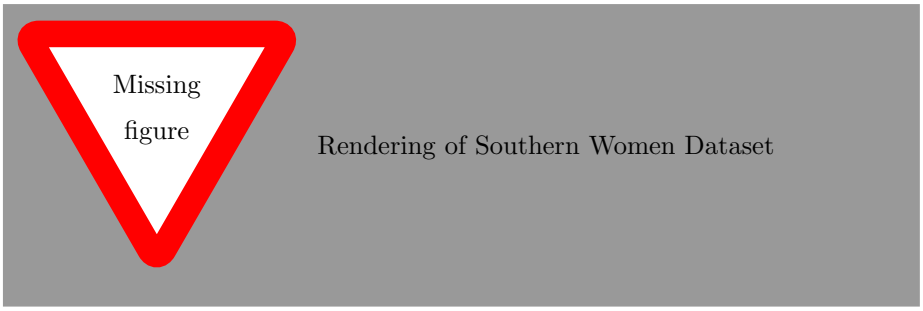


Figure 2: A rendering of the Southern Women Dataset

in which you choose to represent them depends on the question you're trying to answer. For example, one might imagine having two types of nodes in a network. One type of node will represent a person and another type of node will represent an event. An edge is drawn between a person and an event if a person attended a given event and person A is considered connected to person B if they both attended the same event. One such example of this is the *Southern Women Dataset.*[DGG41] This dataset is another example of a community detection problem because after analysis of the data, it was found that women in the group were split into two discrete subgroups.

1.2 Technological Networks

1.3 Information Networks

1.4 Biological Networks

SEC: Definition of a Network

SEC: Different Types of Network

2 Properties of Networks

SEC: Interesting Properties of Networks

3 Community Detection

SEC: Introduction to Community Detection

SEC: Traditional Methods of Community Detection

SEC: Spectral Methods of Community Detection

4 Applications of Community Detection

SEC: Applications of Community Detection

SEC: Figure out an interesting thing to write some of my own code for

References

[DGG41] Allison Davis, Burleigh B. Gardner, and Mary R. Gardner. *Deep South; a Social Anthropological Study of Caste and Class*. The Univ. of Chicago Press, 1941.

[HSS08] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

[Kob12] Stephen G. Kobourov. Spring embedders and force directed graphs. 2012.

[kon17] Zachary karate club network dataset – KONECT, October 2017.

[Kun13] Jérôme Kunegis. KONECT – The Koblenz Network Collection. In *Proc. Int. Conf. on World Wide Web Companion*, pages 1343–1350, 2013.

[New10] Mark Newman. *Networks: An Introduction*. 2010.

[Zac77] Wane W. Zachary. An information flow model for conflict in small groups. 1977.

Is this referencing style appropriate?

Is the information per entry in my bibliography sufficient?