Social Networks 39 (2014) 14–32

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Measuring segregation in social networks

Michał Bojanowski[a,*], Rense Corten[b]

[a] Interdisciplinary Centre for Mathematical and Computational Modelling (ICM), University of Warsaw, Prosta 69, 00-838 Warsaw, Poland
[b] Department of Sociology/ICS, Utrecht Unviersity, Padualaan 14, 3584 CH Utrecht, The Netherlands

ARTICLE INFO

MSC:
91D30

Keywords:
Social networks
Segregation
Homophily
Measurement

ABSTRACT

Network homophily is a pattern in which ties are more likely to exist between nodes similar to each other. It is frequently observed for various types of social relations. At the same time, segregation is often encountered in urban areas as a tendency of families to occupy neighborhoods inhabited by other families similar to them. In this paper we conceptualize both phenomena as in the language of networks of interlinked positions occupied by a population of actors characterized by some node-level attribute. We review existing indexes and approaches to measuring the extent of homophily/segregation in social networks. We pursue an approach of, first, specifying a set of properties that a generic segregation measure might possess, and which, in our view, are relevant in substantial contexts. Second, we check which measures satisfy which properties. The use of measures is illustrated with four empirical examples. Given the particular application and the need for some descriptive measure of segregation, the results presented in this paper can help in selecting an optimal measure for the task at hand. We conclude that the most crucial aspects for the choice of a particular segregation measure include (1) whether the network ties or actors' attributes are assumed to be subject to change, and (2) how one should treat the presence of network isolates.

1. Introduction

In many types of social relations, ties are more likely to form between similar entities than between dissimilar entities. For example, individuals tend to marry others who are similar in terms of age, education, and socio-economic status (Kalmijn, 1998). The discussion of important matters, friendship, and social support also share this feature of *homophily* (see McPherson et al., 2001 for an extensive review of the empirical evidence regarding homophily). We also observe homophily in situations in which individuals affect or influence each other (Erickson, 1988; Cialdini and Goldstein, 2004). For example, people tend to be strongly influenced by others when choosing cultural products (Salganik and Watts, 2009), and friends tend to have similar opinions, especially when the choice of friends is somewhat constrained by the social context (de Klepper et al., 2010).

A related phenomenon, often discussed outside of the social networks literature, is *segregation*. Massey and Denton (1988) defined segregation as "the degree to which two or more groups live separately from one another" in the context of racial segregation of city neighborhoods. The concept of segregation is also applied to the "unequal" distribution of two or more groups of people across different units or social positions. Racial segregation of neighborhoods focuses on the distribution of people belonging to different racial groups across neighborhoods or city blocks constituting the units. In a largely similar fashion, Charles and Grusky (1995) address the way in which groups of men and women are unequally represented in different occupational classes. The literatures on ethnic segregation and gender segregation both emphasize the constraining aspect of segregation as a form of social organization because it places "limits on interactions" (van der Zanden, 1972) and induces a "form of isolation which places limits upon contact, communication, and social relations" (Hunt and Walker, 1974).

While homophily in networks and segregation in neighborhoods or occupations may emerge from very different social processes, the *outcome* in each case is a social structure of interrelated positions occupied by a population of actors consisting of at least two groups. This structure can generally be modeled as a network with the nodes corresponding to the actors and the links corresponding to the relations between the actors. For example, school children from different ethnic groups in a newly assembled class start to form friendships with one another. The typical

* Corresponding author. Tel.: +48 22 8749 427.
E-mail addresses: m.bojanowski@uw.edu.pl (M. Bojanowski), r.corten@uu.nl (R. Corten).
URLs: http://www.bojanorama.pl (M. Bojanowski), http://www.rensecorten.org (R. Corten).

http://dx.doi.org/10.1016/j.socnet.2014.04.001
0378-8733/

outcome of preferential friendship formation processes is a highly homophilous network in which the nodes correspond to children and the links to friendship (Moody, 2001). As another example, consider families of different ethnicities moving to a neighborhood. The neighborhood consists of heterogeneously placed dwellings. In this context, the locations of the dwellings and their spatial proximities can be represented as a network in which the nodes correspond to dwellings and the edges link the dwellings that are adjacent to each other. If the dwellings become occupied by the families, each node of the graph is characterized by the ethnic group of the resident family. Therefore, the outcome is again a network with a node-level attribute designating the *groups* of the nodes. As a third example, consider a contagion-like process, in which some trait spreads in a social network. Also in this case, the outcome is a social structure of types of actors, in this case "infected" and non-infected.

In different literatures, researchers have identified various dimensions of segregation (see Massey and Denton, 1988, for a complete discussion and proposed measures). For the purpose of this paper, we take the view of segregation as (the lack of) *exposure*: the extent to which groups are exposed to one another by occupying nearby positions. This aspect of segregation is intrinsically *relational*, which brings us very close to the social network literature.

Like other types of segregation, segregation in social networks may emerge from different types of processes. Probably the more familiar process is one in which we have a population of actors with fixed attributes (say, gender or ethnicity), among whom a social network is formed. The extent to which actors connect to others with the same attributes generates a level of segregation. However, the same result may also be obtained if the network is fixed and instead the attribute (say, adoption of a certain behavior) change. Segregation then emerges if this attribute tends to cluster in certain parts of the network.

It is important to stress that, while the two processes may substantively be very different, the end result is equivalent from a measurement perspective, namely a situation in which actors with certain attributes are to some degree connected to others with the same attributes. Yet, as will become clear in our analysis, the distinction between the two processes has implications for the interpretation of a given measure of segregation.

A frequent goal of empirical investigations is to compare specific outcomes across different groups, settings, or time points. For example, one could compare different year groups, schools, or classes with respect to the level of friendship segregation (Moody, 2001). In other settings, one might want to compare different districts of a city, or several cities, in terms of the ethnic residential segregation of neighborhoods (Freeman and Sunshine, 1970). Performing such comparisons necessitates the *measurement* of the *level of segregation* in the given network.

Various measures and approaches have been proposed in the network literature. Although these measures are intended for describing the same phenomenon, they originate from different literatures, follow different logics, and are typically proposed without referencing one another. Thus, it is possible for different measures to lead to different conclusions in the same situation. To our knowledge, no systematic overview of the available measures exists. In this paper, we provide a systematic overview of existing segregation measures and highlight the similarities and differences between those measures, with the goal of enabling the researchers to choose the right measure for their respective purposes.

The somewhat dissatisfying state of affairs concerning the measurement of network segregation may be attributed to the same causes that Duncan and Duncan (1955) identified in the realm of segregation measurement (in the stratification sense) in the 50s, namely "naive operationalism" and "[arbitrarily] matching some convenient numerical procedure with the verbal concept of segregation". What is needed is a measurement theory to enable the careful theoretical grounding of segregation measurement.

One particular strategy for building this theoretical basis is the *axiomatic method*. The axiomatic method starts by positing a set of basic properties, or axioms, that a generic measure should possess. In the deductive steps that follow, the goal is to derive classes of measures that logically result from different combinations of the proposed axioms. In the ideal case, the ultimate goal is to arrive at collections of axioms that pin down a single measure of a concept at hand. In other words, given a certain collection of axioms, there exists one and only one measure that simultaneously satisfies all of them.

The axiomatic method has been fruitfully applied in the social sciences. Examples include such diverse domains as utility measurement (Suppes and Winet, 1955), measurement of inequality (Schwartz and Winship, 1980; Cowell and Kuga, 1981; Chakravarty, 1999), income mobility (Cowell, 1985), numerous problems in social choice theory such as the axiomatization of the simple majority rule (May, 1952) or various implications of the assumptions about measurability and comparability of individual utility functions (for example, d'Aspremont and Gevers, 1977, 1985). Regarding segregation, much of the progress in the social stratification research on segregation has been made through the employment of an axiomatic approach (or its associated elements) in the work of James and Tauber (1985), in the later work by Reardon and Firebaugh (2002a) and others (e.g., Egan et al., 1998; Massey and Denton, 1998; Grannis, 2002; Reardon and Firebaugh, 2002b), and recently in work by Alonso-Villar and del Río (2010).

While we believe that a full axiomatization of segregation in social networks is desirable, in this paper we opt for a more practical approach by providing a systematic overview of existing approaches to measuring segregation. We do so by considering a set of atomic properties that a generic segregation measure might possess, and that we believe have practical consequences for research. We then compare existing measures of segregation against this set of properties. Although we do not provide definite results in the form of axiomatizations, we believe that what follows provides an attractive perspective on the problem. The results we obtained should enable researchers to choose an appropriate measure in a particular substantive research context.[1]

In the following section, we define the notation that will be used in the remainder of the paper. In Section 3, we formulate the properties that will guide our analyses of existing segregation measures. Then, the main part of the paper is devoted to an overview and analysis of nine existing segregation measures (Section 4). For each measure, we provide a brief explanation and verify the extent to which the measure conforms to the properties formulated in Section 3. We then demonstrate the use of the measures we discussed by way of a number of empirical examples in Section 5. In the concluding Section 6, we summarize the results of this endeavor and discuss the implications of the results on the practical use of the measures reviewed.

2. Definitions and notation

We introduce the necessary notation and basic definitions that will be used throughout the paper. The notation is loosely based on the standards proposed by Wasserman and Faust (1994).

Network nodes The set of nodes is denoted by $\mathcal{N} = \{1, \ldots, i, \ldots, N\}$.

[1] Instead of "reviewing" the measures, a truly axiomatic method would be to combine the axioms and arrive at some parametrized class(es) of measures. That, however, is beyond the scope of this paper.