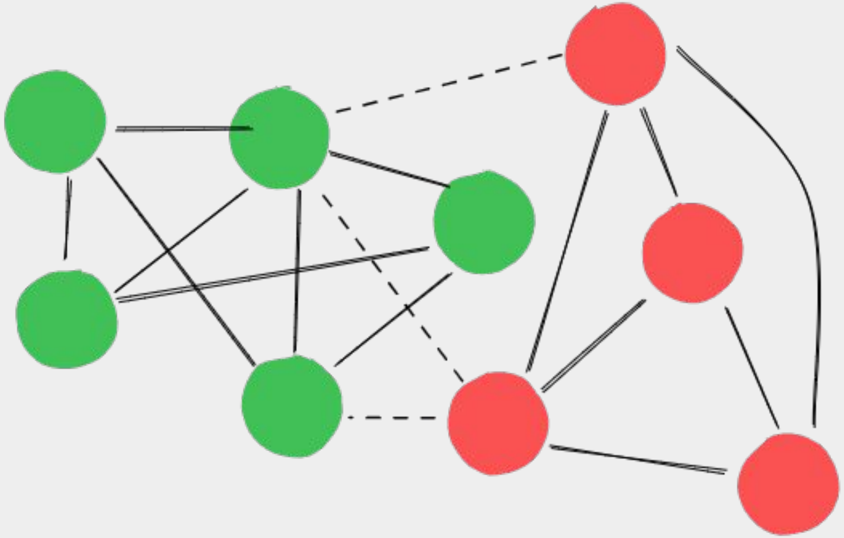
Small Worlds

ivanovitch.silva@ufrn.br
@ivanovitchm

Birds of a Feather



Often, nodes that are connected to each other in a social network tend to be similar in their features. This property is captured by a popular proverb: "birds of a feather flock together." Its technical name is **homophily**. The metric that evaluate homophily is named **assortativity**.

Dating apps leverage this kind of homophily by recommending matches based on shared personality traits.

Echo Chamber

There is a **dark side of homophily**, too. On social media, it is exceedingly easy to connect with people who share our worldviews and unfriend or unfollow people with different opinions — all it takes is a tap of our finger.

Mixing patterns in networks

M. E. J. Newman
Department of Physics, University of Michigan, Ann Arbor, MI 48109–1120 and
Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501

We study assortative mixing in networks, the tendency for vertices in networks to be connected to other vertices that are like (or unlike) them in some way. We consider mixing according to discrete characteristics such as language or race in social networks and scalar characteristics such as age. As a special example of the latter we consider mixing according to vertex degree, i.e., according to the number of connections vertices have to other vertices: do gregarious people tend to associate with other gregarious people? We propose a number of measures of assortative mixing appropriate to the various mixing types, and apply them to a variety of real-world networks, showing that assortative mixing is a pervasive phenomenon found in many networks. We also propose several models of assortatively mixed networks, both analytic ones based on generating function methods, and numerical ones based on Monte Carlo graph generation techniques. We use these models to probe the properties of networks as their level of assortativity is varied. In the particular case of mixing by degree, we find strong variation with assortativity in the connectivity of the network and in the resilience of the network to the removal of vertices.

I. INTRODUCTION

The techniques of statistical physics were developed to study the properties of systems of many interacting particles, atoms, or molecules, but their applicability is wider than this, and recent work has fruitfully applied these techniques to economies, ecosystems, social interactions, the Internet, and many other systems of current interest. The component parts of these systems, the analogs of atoms and molecules, such things as traders in a market, or computers on the Internet, are not usually connected together on a regular lattice as the atoms of a crystal are. Nor indeed do their patterns of connection normally fit any simple low-dimensional structure. Instead they fall on some more generalized "network," which may be more or less random depending on the nature of the system. The broadening of the scope of statistical physics to cover these systems has therefore led us to the consideration of the structure and function of networks, as one of the fundamental steps to understanding real-world phenomena of many kinds. Useful reviews of work in this area can be found in Refs. 1, 2, 3.

Recent studies of network structure have concentrated on a small number of properties that appear to be common to many networks and can be expected to affect the functioning of networked systems in a fundamental way. Among these, perhaps the best studied are the "small-world effect" [4, 5], network transitivity or "clustering" [5], and degree distributions [6, 7]. Many other properties however have been examined and may be equally important, at least in some systems. Examples include resilience to the deletion of network nodes [8, 9, 10, 11, 12], navigability or searchability of networks [13, 14, 15], community structure [16, 17, 18], and spectral properties [19, 20, 21]. In this paper we study another important network feature, the correlations between properties of adjacent network nodes known in the ecology and epidemiology literature as "assortative mixing."

The very simplest representation of a network is a collection of points, usually called vertices or nodes, joined together in pairs by lines, usually called edges or links. More sophisticated network models may introduce other properties of the vertices or the edges. Edges for example may be directed—they point in one particular direction—or may have weights, lengths, or strengths. Vertices can also have weights or other numerical quantities associated with them, or may be drawn from some discrete set of vertex types. In the study of social networks, the patterns of connections between people in a society, it has long been known that edges do not connect vertices regardless of their property or type. Patterns of friendship between individuals for example are strongly affected by the language, race, and age of the individuals in question, among other things. If people prefer to associate with others who are like them, we say that the network shows assortative mixing or assortative matching. If they prefer to associate with those who are different it shows disassortative mixing. Friendship is usually found to be assortative by most characteristics.

Assortative mixing can have a profound effect on the structural properties of a network. For example, assortative mixing of a network by a discrete characteristic will tend to break the network up into separate communities. If people prefer to be friends with others who speak their own language, for example, then one might expect countries with more than one language to separate into communities by language. Assortative mixing by age could cause stratification of societies along age lines. And while the main focus of this paper is on social networks, it is reasonable to suppose that similar mixing effects are seen in non-social networks also. We will give some examples of this in Section III A.

In this paper we study assortative mixing of various types using empirical network data, analytic models, and numerical simulation. We demonstrate that assortative (or disassortative) mixing is indeed present in many networks, show how it can be measured, and examine its

https://arxiv.org/pdf/cond-mat/0209450.pdf

		women				
		black	hispanic	white	other	a_i
men	black	0.258	0.016	0.035	0.013	0.323
	hispanic	0.012	0.157	0.058	0.019	0.247
	white	0.013	0.023	0.306	0.035	0.377
	other	0.005	0.007	0.024	0.016	0.053
	b_i	0.289	0.204	0.423	0.084	

TABLE I: The mixing matrix e_{ij} and the values of a_i and b_i for sexual partnerships in the study of Catania *et al.* [23]. After Morris [24].

$$\sum_{ij} e_{ij} = 1, \quad \sum_j e_{ij} = a_i, \quad \sum_i e_{ij} = b_j$$

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} = \frac{Tr(e) - \|e^2\|}{1 - \|e^2\|}$$

$$r_{min} = -\frac{\sum_i a_i b_i}{1 - \sum_i a_i b_i}$$

```
nx.attribute_assortativity_coefficient(G, "attribute")
nx.attribute_mixing_matrix(G, "attribute")
```

NetworkX
Network Analysis in Python

Fuzzy communities and the concept of bridgeness in complex networks

Tamás Nepusz*
Budapest University of Technology and Economics
Department of Measurement and Information Systems
H-1521 Budapest, P.O.Box 91, Hungary

Andrea Petróczi
Kingston University, School of Life Sciences
Kingston-upon-Thames,
KT1 2EE Surrey, United Kingdom

László Négyessy
Neurobionics Research Group
Hungarian Academy of Sciences - Péter Pázmány Catholic University - Semmelweis University
H-1094 Budapest, Tűzoltó u. 58., Hungary

Fülöp Bazsó†
KFKI Research Institute for Particle and Nuclear Physics of the Hungarian Academy of Sciences
Department of Biophysics
H-1525 Budapest, P.O.Box 49, Hungary

(Dated: May 31, 2018)

We consider the problem of fuzzy community detection in networks, which complements and expands the concept of overlapping community structure. Our approach allows each vertex of the graph to belong to multiple communities at the same time, determined by exact numerical membership degrees, even in the presence of uncertainty in the data being analyzed. We created an algorithm for determining the optimal membership degrees with respect to a given goal function. Based on the membership degrees, we introduce a new measure that is able to identify outlier vertices that do not belong to any of the communities, bridge vertices that belong significantly to more than one single community, and regular vertices that fundamentally restrict their interactions within their own community, while also being able to quantify the centrality of a vertex with respect to its dominant community. The method can also be used for prediction in case of uncertainty in the dataset analyzed. The number of communities can be given in advance, or determined by the algorithm itself using a fuzzified variant of the modularity function. The technique is able to discover the fuzzy community structure of different real world networks including, but not limited to social networks, scientific collaboration networks and cortical networks with high confidence.

I. INTRODUCTION

Recent studies revealed that graph models of many real world phenomena exhibit an overlapping community structure, which is hard to grasp with the classical graph clustering methods where every vertex of the graph belongs to exactly one community [1]. This is especially true for social networks, where it is not uncommon that individuals in the network belong to more than one community at the same time. Individuals who connect groups in the network function as "bridges", hence the concept of "bridge" is defined as the vertices that cross structural holes between discrete groups of people [2]. It is therefore important to define a quantity that measures the commitment of a node to several communities in order to obtain a more realistic view of these networks.

The intuitive meaning of a bridge vertex may differ in different kinds of networks that exist beyond sociometrics. In protein interaction networks, bridges can be proteins with multiple roles. In cortical networks containing brain areas responsible for different modalities (for instance, visual and tactile input processing), the bridges are presumably the areas that take part in the integration and higher level processing of sensory signals. In word association networks, words with multiple meanings are likely to be bridges [35]. The state-of-the-art overlapping community detection algorithms [1, 3, 4, 5] are not able to quantify the notion of bridgeness, while other attempts at quantifying it (e.g., the participation index [6]) are only concerned with non-overlapping communities.

To emphasize the importance of bridge vertices in community detection and to illustrate the concept, we take a simple graph shown on Fig. 1(a) as an example. A visual inspection of this graph most likely suggests two

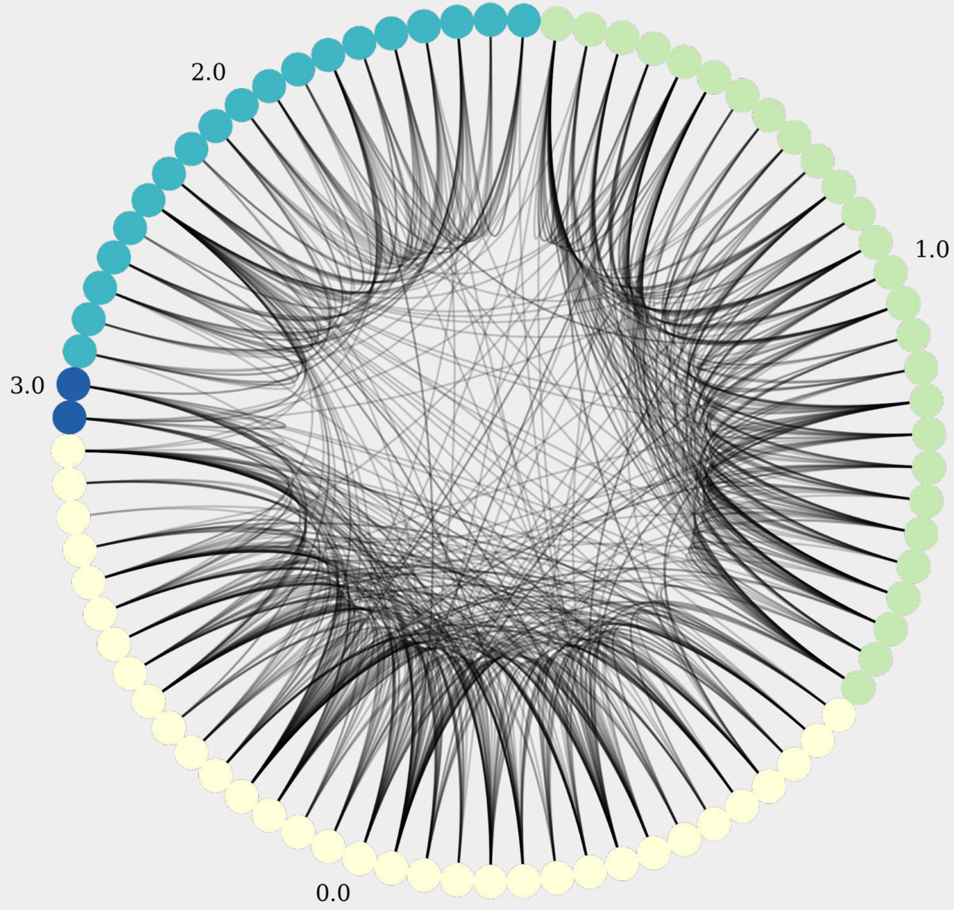
*Electronic address: nepusz@mit.bme.hu; Also at KFKI Research Institute for Particle and Nuclear Physics of the Hungarian Academy of Sciences, Department of Biophysics, Budapest, Hungary and Kingston University, School of Life Sciences, Kingston-upon-Thames, United Kingdom.
†Electronic address: bazso@sunserv.kfki.hu; Also at Polytechnical Engineering College Subotica, Marka Oreškovića 16, 24000 Subotica, Serbia.

https://arxiv.org/pdf/0707.1646.pdf

Friendship network of a UK university faculty

The personal friendship network of a faculty of a UK university, consisting of **81 vertices** (individuals) and **817 directed and weighted connections**. The school affiliation of each individual is stored as a vertex attribute. This dataset can serve as a testbed for community detection algorithms.

http://hal.elte.hu/~nepusz/research/datasets/




```
G = nx.read_graphml('univ_dataset_TSPE.graphml')

G.nodes(data=True)
{'n0': {'group': 2.0, 'id': 'n0'} ......}

nx.attribute_assortativity_coefficient(G,"group")
0.7053802318353712
```

```
nx.attribute_mixing_matrix(G,'group')

array([[0.3880049 , 0.0501836 , 0.01591187, 0.01713586],
       [0.02937576, 0.30599755, 0.00734394, 0.00244798],
       [0.02570379, 0.01591187, 0.11750306, 0.00244798],
       [0.01346389, 0.00367197, 0.00244798, 0.00244798]
])
```

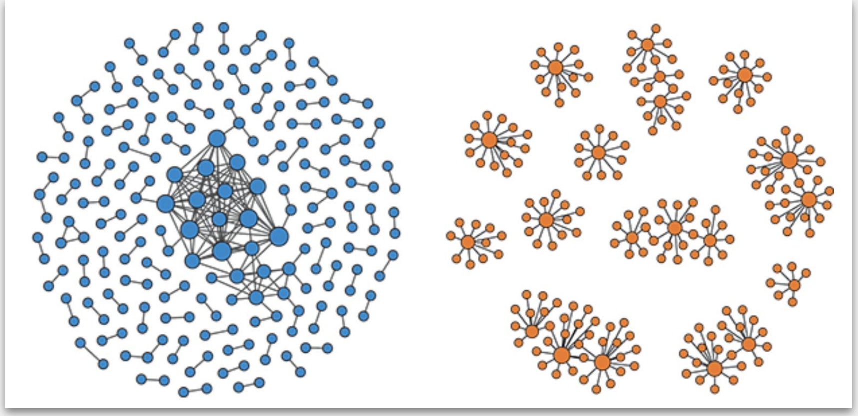
Measuring discrete assortative mixing

NetworkX
Network Analysis in Python

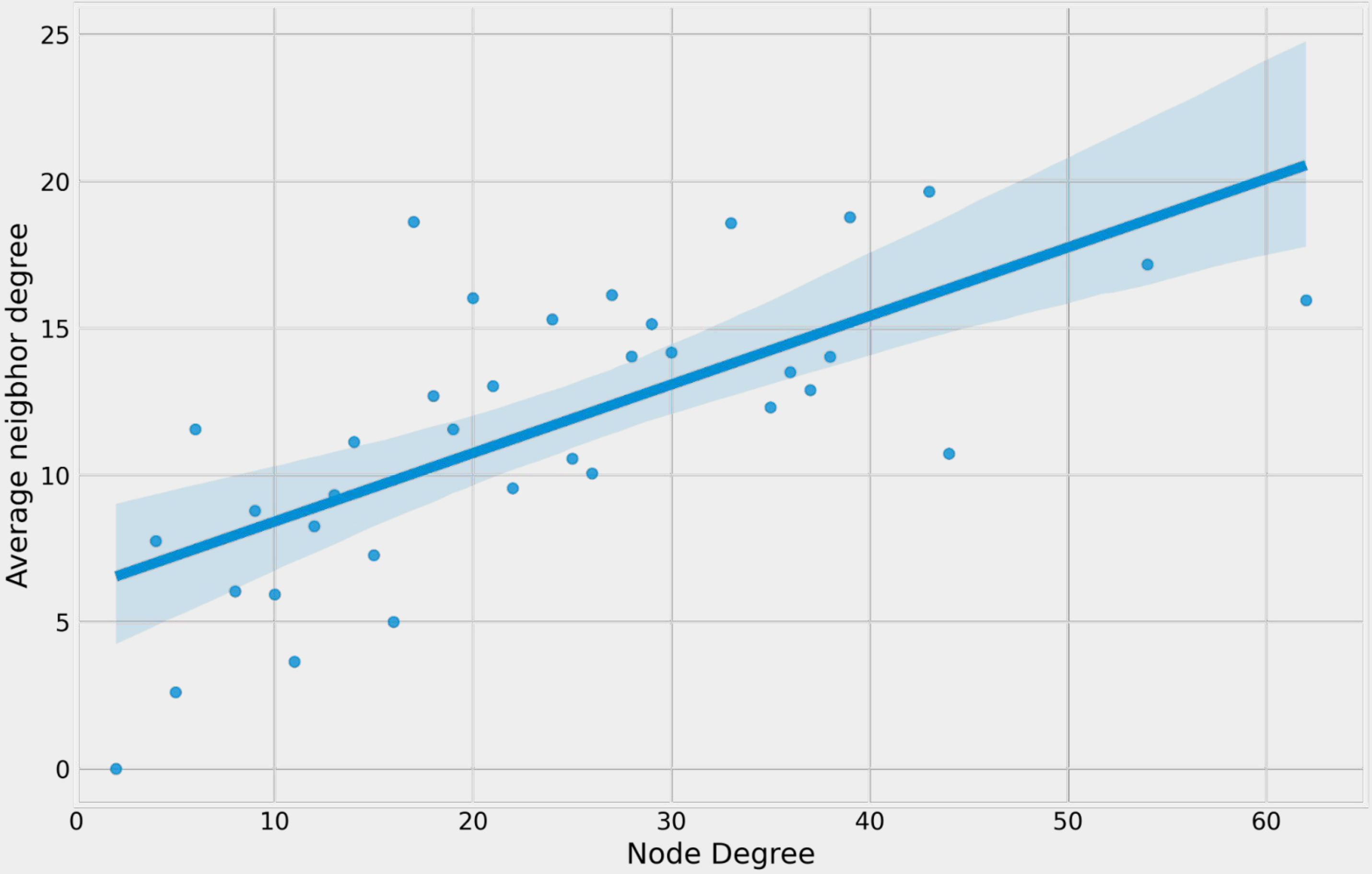
How about the Assortativity related to Degree?

Assortative network

Disassortative network



Assortativity is not exclusive of social networks; nodes in many types of networks have properties that may be similar among neighbors. For example, nodes in any network have the fundamental property of **degree**. Assortativity based on degree is called **degree assortativity** or **degree correlation**.




```
degree, avg_neigh_degree = zip(*nx.average_degree_connectivity(G).items())
```

NetworkX
Network Analysis in Python

arXiv:cond-mat/0209450v2 [cond-mat.stat-mech] 4 Feb 2003

Mixing patterns in networks

M. E. J. Newman
Department of Physics, University of Michigan, Ann Arbor, MI 48109–1120 and
Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501

We study assortative mixing in networks, the tendency for vertices in networks to be connected to other vertices that are like (or unlike) them in some way. We consider mixing according to discrete characteristics such as language or race in social networks and scalar characteristics such as age. As a special example of the latter we consider mixing according to vertex degree, i.e., according to the number of connections vertices have to other vertices: do gregarious people tend to associate with other gregarious people? We propose a number of measures of assortative mixing appropriate to the various mixing types, and apply them to a variety of real-world networks, showing that assortative mixing is a pervasive phenomenon found in many networks. We also propose several models of assortatively mixed networks, both analytic ones based on generating function methods, and numerical ones based on Monte Carlo graph generation techniques. We use these models to probe the properties of networks as their level of assortativity is varied. In the particular case of mixing by degree, we find strong variation with assortativity in the connectivity of the network and in the resilience of the network to the removal of vertices.

I. INTRODUCTION

The techniques of statistical physics were developed to study the properties of systems of many interacting particles, atoms, or molecules, but their applicability is wider than this, and recent work has fruitfully applied these techniques to economies, ecosystems, social interactions, the Internet, and many other systems of current interest. The component parts of these systems, the analogs of atoms and molecules, such things as traders in a market, or computers on the Internet, are not usually connected together on a regular lattice as the atoms of a crystal are. Nor indeed do their patterns of connection normally fit any simple low-dimensional structure. Instead they fall on some more generalized "network," which may be more or less random depending on the nature of the system. The broadening of the scope of statistical physics to cover these systems has therefore led us to the consideration of the structure and function of networks, as one of the fundamental steps to understanding real-world phenomena of many kinds. Useful reviews of work in this area can be found in Refs. 1, 2, 3.

Recent studies of network structure have concentrated on a small number of properties that appear to be common to many networks and can be expected to affect the functioning of networked systems in a fundamental way. Among these, perhaps the best studied are the "small-world effect" [4, 5], network transitivity or "clustering" [5], and degree distributions [6, 7]. Many other properties however have been examined and may be equally important, at least in some systems. Examples include resilience to the deletion of network nodes [8, 9, 10, 11, 12], navigability or searchability of networks [13, 14, 15], community structure [16, 17, 18], and spectral properties [19, 20, 21]. In this paper we study another important network feature, the correlations between properties of adjacent network nodes known in the ecology and epidemiology literature as "assortative mixing."

The very simplest representation of a network is a collection of points, usually called vertices or nodes, joined together in pairs by lines, usually called edges or links. More sophisticated network models may introduce other properties of the vertices or the edges. Edges for example may be directed—they point in one particular direction—or may have weights, lengths, or strengths. Vertices can also have weights or other numerical quantities associated with them, or may be drawn from some discrete set of vertex types. In the study of social networks, the patterns of connections between people in a society, it has long been known that edges do not connect vertices regardless of their property or type. Patterns of friendship between individuals for example are strongly affected by the language, race, and age of the individuals in question, among other things. If people prefer to associate with others who are like them, we say that the network shows assortative mixing or assortative matching. If they prefer to associate with those who are different it shows disassortative mixing. Friendship is usually found to be assortative by most characteristics.

Assortative mixing can have a profound effect on the structural properties of a network. For example, assortative mixing of a network by a discrete characteristic will tend to break the network up into separate communities. If people prefer to be friends with others who speak their own language, for example, then one might expect countries with more than one language to separate into communities by language. Assortative mixing by age could cause stratification of societies along age lines. And while the main focus of this paper is on social networks, it is reasonable to suppose that similar mixing effects are seen in non-social networks also. We will give some examples of this in Section III A.

In this paper we study assortative mixing of various types using empirical network data, analytic models, and numerical simulation. We demonstrate that assortative (or disassortative) mixing is indeed present in many networks, show how it can be measured, and examine its

https://arxiv.org/pdf/cond-mat/0209450.pdf

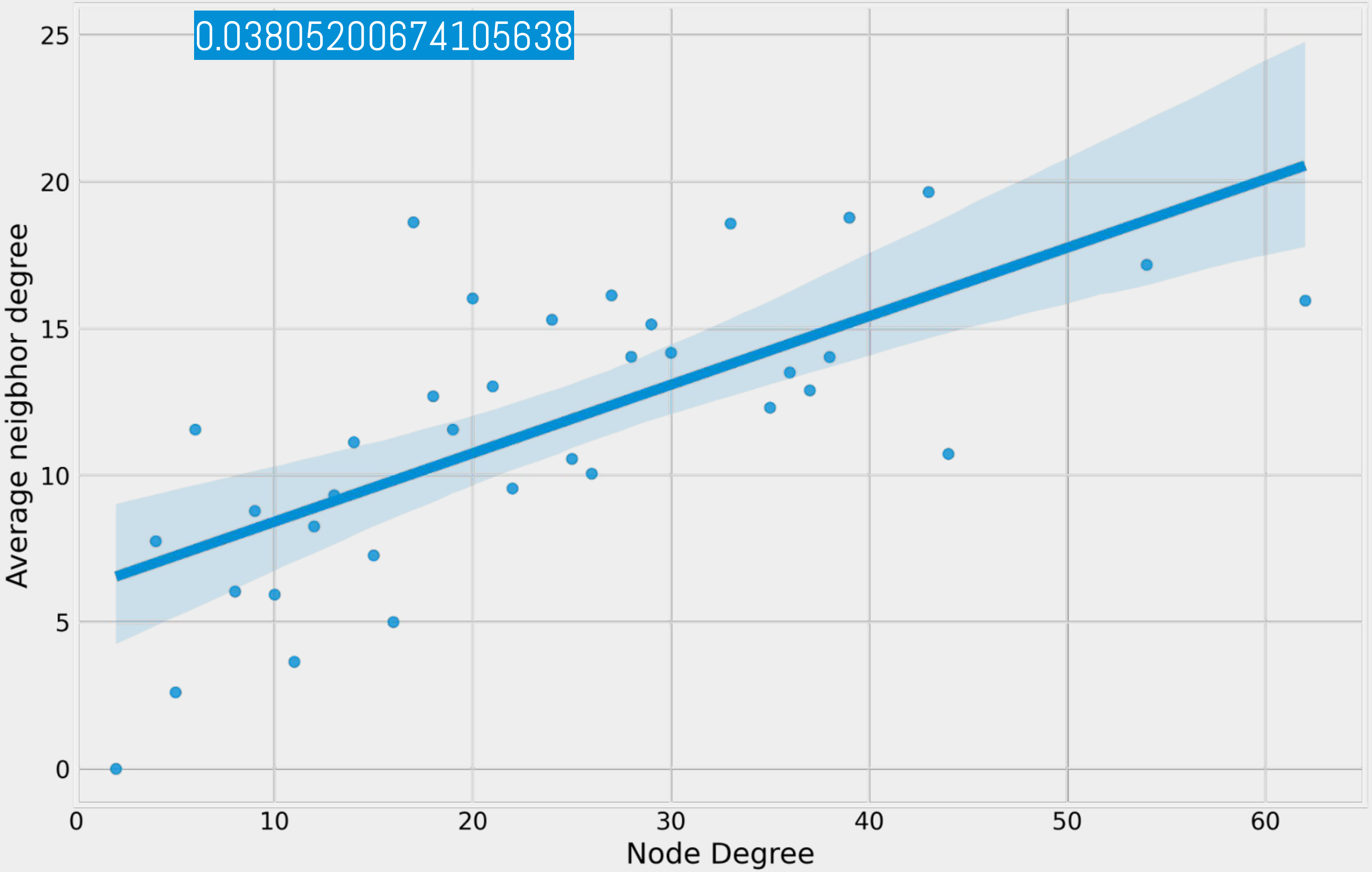
The degree correlation coefficient

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_q^2} \qquad -1 \le r \le 1$$

$$\begin{array}{cc}
 & \begin{array}{ccc} k_1 & k_2 & k_3 \end{array} \\
\begin{array}{c} j_1 \\ j_2 \\ j_3 \end{array} & \left(\begin{array}{ccc} 0.073265410 & 0.12291599 & 0.04662765 \\ 0.06307208 & 0.18616924 & 0.09045399 \\ 0.153551780 & 0.15433145 & 0.1096124100 \end{array}\right) \\
q' & \begin{array}{ccc} 0.28988927 & 0.46341668 & 0.24669405 \end{array}
\end{array}
\begin{array}{c} q \\ \rightarrow 0.24280905 \\ \rightarrow 0.33969531 \\ \rightarrow 0.41749564 \end{array}$$

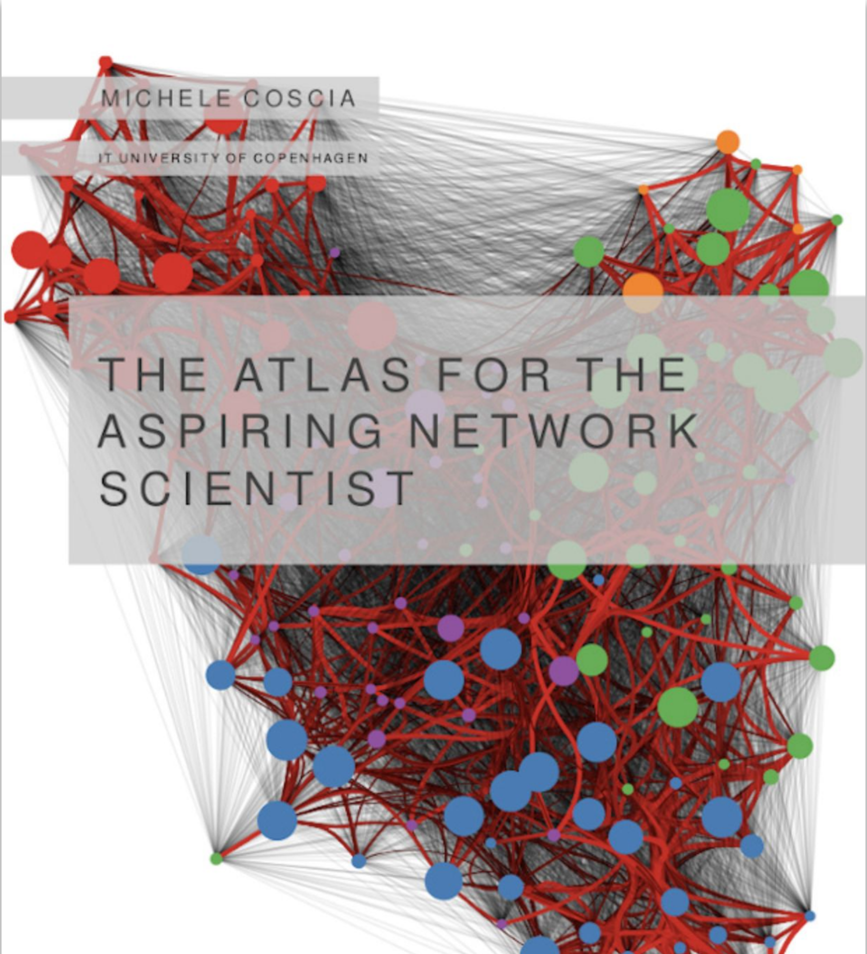
std(q) = 0.071

std(q') = 0.093




```
nx.degree_assortativity_coefficient(G)
```

NetworkX
Network Analysis in Python



https://www.networkatlas.eu/

Further Reading

Chapter 26 Homophily
Chapter 27 Quantitative Assortativity