Network heterogeneity in an undirected network

Hu, X.; Leydesdorff, L.; Rousseau, R.

Published in:
21st International Conference on Science and Technology Indicators: book of proceedings

DOI:
10.4995/STI2016.2016.4543

Creative Commons License (see https://creativecommons.org/use-remix/cc-licenses):
CC BY-NC-ND

Citation for published version (APA):
BOOK OF PROCEEDINGS

STI Conference 2016 · València
Peripheries, frontiers and beyond
14 · 16 September 2016
Universitat Politècnica de València

www.sti2016.org
Network heterogeneity in an undirected network

Xiaojun Hu,* Loet Leydesdorff,** and Ronald Rousseau***

* xjhu@zju.edu.cn
Medical Information Center, Zhejiang University School of Medicine, Hangzhou 310058 (China)

** loet@leydesdorff.net
University of Amsterdam, Amsterdam School of Communication Research (ASCoR), PO Box 15793, 1001 NG Amsterdam (The Netherlands)

*** ronald.rousseau@kuleuven.be
KU Leuven, Dept. of Mathematics, Celestijnenlaan 200B, B-3001 Leuven (Belgium) & University of Antwerp (UA), IBW, B-2000 Antwerp (Belgium)

ABSTRACT
Networks may be homogeneous in the sense that all nodes are of the same type or they may be heterogeneous in the sense of containing many types of nodes. Once a choice about the types of nodes to be distinguished is made, it is still an open question how to define an acceptable measure of heterogeneity. A proposal to solve this problem is provided in this contribution. In our view the term heterogeneity in a network implies that links between nodes of different type are gauged positively, while links between nodes of the same type should not contribute to a heterogeneity value. Hence a high heterogeneity value refers to a tightly woven net between dissimilar things. We value this property so high that networks without external links receive the same heterogeneity value as homogeneous networks. Concretely, units that determine the value of a heterogeneity measure are links between nodes of different types. These considerations lead to a new measure for heterogeneity derived from a true diversity measure. We claim that we are now able to measure the heterogeneity of networks in a much more precise way than was possible before. An example related to interdisciplinarity is provided. As heterogeneous networks are ubiquitous in the real world, such as in molecular networks, disease networks and trade networks our approach has universal applicability.

INTRODUCTION
Networks may be homogeneous in the sense that all nodes are of the same type or they may contain many types of nodes. A bipartite network linking authors with papers provides an example. Similarly, the nodes in a network of bibliographically coupled or co-cited articles can be attributed to the journal in which they are published.

In this contribution the term network heterogeneity refers to links connecting different types of nodes. Whether two nodes are considered to be of different type depends on the application one has in mind. Once a choice about the types of nodes to be distinguished is made, it is still an open question how to define an acceptable measure of heterogeneity. In this contribution, we propose a solution to this problem for the case of undirected networks. Links between nodes of the same type are called internal links, while links between nodes of different type are called external links. Heterogeneous networks are ubiquitous in the real world; hence a

1 This work was supported by the National Natural Science Foundation of China (NSFC grant No. 71573225).

This work is licensed under a Creative Commons License: Attribution-NonCommercial-NoDerivatives 4.0 International
precise definition of heterogeneity is required. In this contribution such a measure is proposed.

HETEROGENEOUS NETWORKS AND AN INTERPRETATION AS LAYERED NETWORKS

In our view the term heterogeneity in a network implies that external links are gauged positively, while internal links should not contribute. Moreover, we prefer evenly distributed external links above unevenly distributed ones (Hill, 1973). This is a choice we made, admitting that other choices, depending on the application one has in mind can be made. In this contribution a high heterogeneity value refers to a tightly woven net between dissimilar things. Networks without external links receive the same heterogeneity value as homogeneous networks. Indeed, when different types of nodes are never connected the network is actually a disjoint union of homogeneous networks. Concretely, units that will determine the value of a heterogeneity measure are links between nodes of different types. We stress the point that units are not nodes but external links.

Interpretation as a layered network

If there are N types of nodes these may be considered as belonging to N layers, leading to a layered network. Links connecting different types of nodes, i.e. external links, then become links between different layers. In this way the theory of layered networks (Boccaletti et al., 2014) can be applied to heterogeneous networks. The following example (Fig.1) illustrates how a heterogeneous network can be seen as a layered one.

**Figure 1**: a Representation in a single plane; b Representation as a layered network
DIVERSITY AS MEASURED THROUGH VARIETY AND BALANCE
We claim that node heterogeneity can be considered as a particular form of diversity, in which links between different types of nodes are the essential constituents. Before proposing a measure for this type of heterogeneity we recall that traditionally diversity is measured through the notions of variety and balance (Magurran, 2003).

The definitions of variety and balance
Variety is the number of non-empty categories to which system elements are assigned. Assuming all other things equal, the greater the variety, the greater the diversity. Balance is a function of the assignment of elements across categories. It answers the question: What is the relative number of items of each type? All else being equal, the more balanced the distribution, the larger the diversity. Variety is a positive natural number as categories are numbered in sequence; balance is a function of fractions summing up to one.

HETEROGENEITY
From now on we use the term heterogeneity when referring to the diversity of links between layers of different types of nodes. Moreover, we follow Hill (1973) and Jost (2006, 2009) requiring that heterogeneity measures should be so-called “true” heterogeneity measures. The main point about these measures is that only when working with true measures it makes sense to discuss heterogeneity in terms of ratios or percentages.

Following Jost’s (Jost, 2006) arguments in favor of true diversity we apply his formula qD for the measurement of heterogeneity of networks choosing q = 2. Assuming there are N layers this leads to a Hirschman-Simpson-Herfindahl type of heterogeneity measure, which we denote by HE. If the network itself is denoted by X, we have:

\[ HE(X) = \left( \frac{1}{\sum_{i=1}^{N} \sum_{j=i+1}^{N} p_{ij}^{2}} \right)^{1/2} = \frac{1}{\sum_{i=1}^{N} \sum_{j=i+1}^{N} p_{ij}^{2}} \]

The symbols pij refer to the relative number of links, among external links, between nodes of type i and of type j. In the case of a homogeneous network X - this means that all nodes are of the same type and hence N = 1 - we set HE(X) equal to zero. Similarly, the heterogeneity value of a network without external links is also set equal to zero.

A simple example
We determine the heterogeneity of the example network shown in Fig.1. In this case N=3. Relative proportions are: pI,II = 0.5, pI,III=0.25 and pII,III = 0.25. Hence

\[ HE(X) = \frac{1}{p_{I,II}^2 + p_{I,III}^2 + p_{II,III}^2} = \frac{1}{\frac{1}{4} + \frac{1}{16} + \frac{1}{16}} = \frac{1}{\frac{6}{16}} = \frac{8}{3} \]

What happens when a peripheral node is added?
By a peripheral node we mean a node which forms a type on its own and is linked to no other or exactly one other node. If a node which forms a type on its own is not linked to any other node, then the heterogeneity measure stays the same.
If a node which forms a type on its own is linked to exactly one other node, then N becomes N+1 and the number of different types of links (M) increases by one, too. We denote the original network by X and the one with one extra node attached by X’. If pij = nij/M (where nij is the number of links between nodes of type i and nodes of type j in network X) then in X’ the corresponding pij’ becomes nij/M+1 = pij (M/M+1). Now,

\[ HE(X') = \frac{1}{\sum_{i=1}^{N} \sum_{j=i+1}^{N} (p_{ij})^2 \left( \frac{1}{M+1} \right)^2 + \frac{1}{\sum_{i=1}^{N} \sum_{j=i+1}^{N} (p_{ij})^2 \left( \frac{1}{M+1} \right)^2}} = \frac{(M+1)^2}{\left( \sum_{i=1}^{N} (p_{ij})^2 \left( \frac{1}{M} \right)^2 \right)^{-1}} = \frac{1}{\left( \sum_{i=1}^{N} (p_{ij})^2 \left( \frac{1}{M} \right)^2 \right)^{-1} \left( HE(X) \right)^{-1} + \frac{1}{M^2}} = \frac{(M+1)^2}{\left( \sum_{i=1}^{N} (p_{ij})^2 \left( \frac{1}{M} \right)^2 \right)^{-1}} = \frac{1}{\left( \sum_{i=1}^{N} (p_{ij})^2 \left( \frac{1}{M} \right)^2 \right)^{-1}} = \frac{1}{\left( \sum_{i=1}^{N} (p_{ij})^2 \left( \frac{1}{M} \right)^2 \right)^{-1}} \left( HE(X) \right)^{-1} + \frac{1}{M^2} \left( (HE(X))^{-1} + \frac{1}{M^2} \right)^{-1} \left( \sum_{i=1}^{N} (p_{ij})^2 \left( \frac{1}{M} \right)^2 \right)^{-1}}

EXTENSION TO WEIGHTED LINKS AND TO THE CASE THAT A NODE MAY BE OF DIFFERENT TYPES

When links are valued or weighted positively, formula (1) can still be used, but the pij become relative weights with respect to the total weight. If a node can be of more than one type it is assumed to belong to several different layers, linked to replica of itself. As space does not permit we omit the technical details for this case.

A REAL-WORLD EXAMPLE

As an illustration we determined the heterogeneity of the reference lists of twelve articles which were studied in an earlier article by Rafols and Meyer (2010) in the context of a study on interdisciplinarity of nanobioscience. We refer to the original publications for bibliographic details of these twelve publications. Here we denote them in the same way as in (Rafols & Meyer, 2010). Table 1 provides a comparison between some indicators and the heterogeneity measure (HE) for these 12 papers. Data for the Stirling index (Stirling, 2007) and the mean linkage strength (S) are taken from (Rafols & Meyer, 2010). Results for the Stirling index are obtained based on the distribution of WoS subject categories of references.

Table 1. A comparison between some indicators and our heterogeneity measure

<table>
<thead>
<tr>
<th>First author of article (year of publication)</th>
<th>#WoS fields</th>
<th>Stirling index (refs of refs)</th>
<th>S: mean linkage strength of bc network</th>
<th>HE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burgess (2003)</td>
<td>8</td>
<td>0.14</td>
<td>0.050</td>
<td>3.93</td>
</tr>
<tr>
<td>Funatsu (1995)</td>
<td>6</td>
<td>0.27</td>
<td>0.054</td>
<td>2.24</td>
</tr>
<tr>
<td>Noji (1997)</td>
<td>4</td>
<td>0.15</td>
<td>0.024</td>
<td>5.31</td>
</tr>
<tr>
<td>Ishijima (1998)</td>
<td>7</td>
<td>0.18</td>
<td>0.042</td>
<td>7.41</td>
</tr>
<tr>
<td>Kikkawa (2001)</td>
<td>8</td>
<td>0.16</td>
<td>0.072</td>
<td>5.45</td>
</tr>
<tr>
<td>Kojima (1997)</td>
<td>4</td>
<td>0.24</td>
<td>0.074</td>
<td>4.39</td>
</tr>
<tr>
<td>Okada (1999)</td>
<td>4</td>
<td>0.15</td>
<td>0.107</td>
<td>3.13</td>
</tr>
<tr>
<td>Sakakibara (1999)</td>
<td>6</td>
<td>0.16</td>
<td>0.029</td>
<td>5.29</td>
</tr>
<tr>
<td>Tomishige (2000)</td>
<td>7</td>
<td>0.14</td>
<td>0.104</td>
<td>5.70</td>
</tr>
<tr>
<td>Tomishige (2002)</td>
<td>5</td>
<td>0.15</td>
<td>0.113</td>
<td>4.97</td>
</tr>
<tr>
<td>Yasuda (1998)</td>
<td>4</td>
<td>0.14</td>
<td>0.039</td>
<td>3.98</td>
</tr>
<tr>
<td>Yildiz (2004)</td>
<td>11</td>
<td>0.17</td>
<td>0.065</td>
<td>13.87</td>
</tr>
</tbody>
</table>

This work is licensed under a Creative Commons License: Attribution-NonCommercial-NoDerivatives 4.0 International.
Table 2. Pearson correlation between diversity, network coherence, variety of subject categories and heterogeneity

<table>
<thead>
<tr>
<th></th>
<th>#WoS fields</th>
<th>Stirling</th>
<th>Coherence</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td># WoS fields</td>
<td>1.00</td>
<td>-0.08</td>
<td>-0.02</td>
<td>0.73</td>
</tr>
<tr>
<td>Stirling</td>
<td>1.00</td>
<td>-0.09</td>
<td>-0.17</td>
<td></td>
</tr>
<tr>
<td>Coherence</td>
<td></td>
<td>1.00</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>Heterogeneity</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

It can rightly be argued that the distribution of references of references is an indicator of interdisciplinarity, yet it is not an indicator of WoS category heterogeneity of references. As such we are not surprised that there does not seem to be a relation between the Stirling index and the HE-measure. It seems though that the number of WoS fields present in the reference list plays a significant role. This is to be expected as in this example ‘heterogeneity’ is characterized by the presence of different WoS fields in the reference list. Yet, even this small example shows that there is no simple one-to-one correspondence between the number of WoS categories present and the HE-value.

DISCUSSION AND CONCLUSION
The practical example we elaborated is related to interdisciplinarity, considered here as a special case of heterogeneity. Yet we stress the point that our approach deals with heterogeneity in general and is not restricted to interdisciplinarity. Heterogeneous networks, composed of different types of objects are ubiquitous in the real world. In other words, our new indicator for heterogeneity has universal applicability.

REFERENCES


