Derived networks and multi-mode network analysis

Vladimir Batagelj

IMFM Ljubljana and IAM UP Koper

ARS'17 International Workshop

Challenges in Social network research

Naples, 16-17. May 2017

Outline

Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si

Version (May 17, 2017, 10 : 00): ARS2017vb.pdf

Linked / multi-mode networks

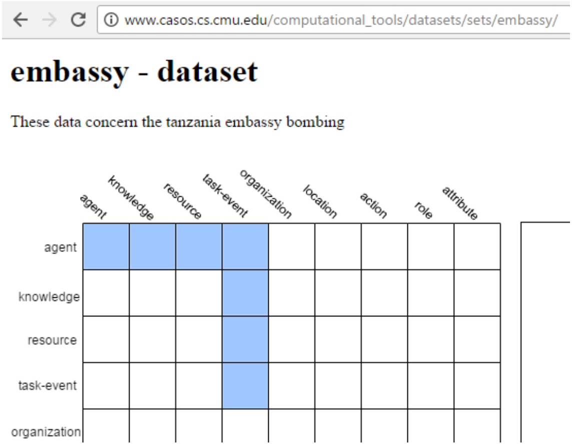
Linked or multi-mode networks are collections of networks over at least two sets of nodes (modes) and consist of some one-mode networks and some two-mode networks linking different modes.

For example: modes are Persons and Organizations. Two one-mode networks describe collaboration among Persons and among Organizations. The linking two-mode network describes membership of Persons to different Organizations.

Linked networks were introduced as a Meta-Matrix approach by Krackhardt and Carley in 1998 [5, 3].

Carley and Diesner

Meta-Matrix Entities	Agent	Knowledge	Resources	Tasks/ Event	Organizations	Location
Agent	Social network	Knowledge network	Capabilities network	Assignment network	Membership network	Agent location network
Knowledge		Information network	Training network	Knowledge requirement network	Organizational knowledge network	Knowledge location network
Resources			Resource network	Resource requirement Network	Organizational Capability network	Resource location network
Tasks/ Events				Precedence network	Organizational assignment network	Task/Event location network
Organizations					Inter-organizational network	Organizatio nal location network
Location						Proximity network


www.casos.cs.cmu.edu/computational_tools/datasets/sets/embassy/
embassy - dataset
These data concern the tanzania embassy bombing
agent
knowledge
resource
task-event
organization
location
action
role
attribute
agent
knowledge
resource
task-event
organization

Analysis of linked networks

We can analyze each network separately using available methods for analysis of one-mode and two-mode networks.

For analysis of linked networks we can also use the constrained blockmodeling approach (Žiberna [10]).

In this presentation we will discuss some issues related to the use of *derived networks* obtained by network multiplication.

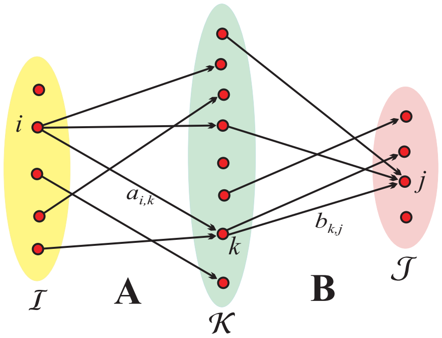
Multiplication of networks

Given a pair of compatible networks $\mathcal{N}_A = ((\mathcal{I}, \mathcal{K}), \mathcal{A}_A, w_A)$ and $\mathcal{N}_B = ((\mathcal{K}, \mathcal{J}), \mathcal{A}_B, w_B)$ with corresponding matrices $\mathbf{A}_{\mathcal{I}\times\mathcal{K}}$ and $\mathbf{B}_{\mathcal{K}\times\mathcal{J}}$ we call a *product of networks* $\mathcal{N}_A$ and $\mathcal{N}_B$ a network $\mathcal{N}_C = ((\mathcal{I}, \mathcal{J}), \mathcal{A}_C, w_C)$, where $\mathcal{A}_C = \{(i,j) : i \in \mathcal{I}, j \in \mathcal{J}, c_{i,j} \neq 0\}$ and $w_C(i,j) = c_{i,j}$ for $(i,j) \in \mathcal{A}_C$. The product matrix $\mathbf{C} = [c_{i,j}]_{\mathcal{I}\times\mathcal{J}} = \mathbf{A} * \mathbf{B}$ is defined in the standard way

$$c_{i,j} = \sum_{k\in\mathcal{K}} a_{i,k} \cdot b_{k,j}$$

In the case when $\mathcal{I} = \mathcal{K} = \mathcal{J}$ we are dealing with ordinary one-mode networks (with square matrices).

Multiplication of networks



$$c_{i,j} = \sum_{k \in N_A(i) \cap N_B^-(j)} a_{i,k} \cdot b_{k,j}$$

If all weights in networks $\mathcal{N}_A$ and $\mathcal{N}_B$ are equal to 1 the value of $c_{i,j}$ counts the number of ways we can go from $i \in \mathcal{I}$ to $j \in \mathcal{J}$ passing through $\mathcal{K}$: $c_{i,j} = |N_A(i) \cap N_B^-(j)|$.

Multiplication of networks

The standard matrix multiplication is too slow to be used for large networks. For sparse large networks we can multiply much faster considering only nonzero elements. In general the multiplication of large sparse networks is a 'dangerous' operation since the result can 'explode' – it is not sparse.

If for the sparse networks $\mathcal{N}_A$ and $\mathcal{N}_B$ there are in $\mathcal{K}$ only few nodes with large degree and no one among them with large degree in both networks then also the resulting product network $\mathcal{N}_C$ is sparse.

The multiplication transforms two linked networks on sets $\mathcal{I} \times \mathcal{K}$ and $\mathcal{K} \times \mathcal{J}$ into a network on the sets $\mathcal{I} \times \mathcal{J}$. Such networks are called *derived* networks. They are usually weighted.

Often we transform a two-mode network $\mathcal{N} = ((\mathcal{U}, \mathcal{V}), \mathcal{L}, w)$ into an ordinary (one-mode) network $\mathcal{N}_r = (\mathcal{U}, \mathcal{A}_r, w_r)$ or/and $\mathcal{N}_c = (\mathcal{V}, \mathcal{A}_c, w_c)$, where $\mathcal{A}_r$ and w_r are determined by the matrix $\mathbf{W}_r = \mathbf{W}\mathbf{W}^T$.

The network $\mathcal{N}_c$ is determined in a similar way by the matrix $\mathbf{W}_c = \mathbf{W}^T\mathbf{W}$.

The networks $\mathcal{N}_r$ and $\mathcal{N}_c$ are analyzed using standard methods.

Bibliographic networks

From a bibliography on selected topic we can construct some two-mode networks:

works $\times$ authors (**WA**),
works $\times$ keywords (**WK**);
works $\times$ journals/publishers (**WJ**);
works $\times$ classification (**WC**)
authors $\times$ institutions (**AI**);
institutions $\times$ countries (states) (**IS**);

and sometimes also the one-mode citation network

works $\times$ works (**Ci**);

where works include papers, reports, books, patents etc.

Besides this we get also at least the partition of works by the publication year, and the vector of number of pages.

First co-authorship network

Let **WA** be the works $\times$ authors two-mode authorship network; $wa_{pi} = 1$ is describing the authorship of author i of work p.

$$\forall p \in W : \sum_{i \in A} wa_{pi} = \text{outdeg}_{WA}(p) = \# \text{ authors of work } p$$

Transposition $\mathcal{N}^T$ or $t(\mathcal{N})$ is a network obtained from $\mathcal{N}$ in which the node sets are interchanged and to all arcs their direction is reversed. $\mathbf{AW} = \mathbf{WA}^T$, $\mathbf{KW} = \mathbf{WK}^T$, ...

The first *co-authorship* network is defined as $\mathbf{Co} = \mathbf{AW} * \mathbf{WA}$

$$co_{ij} = \sum_{p \in W} wa_{pi} wa_{pj} = \sum_{p \in N^-(i) \cap N^-(j)} 1$$

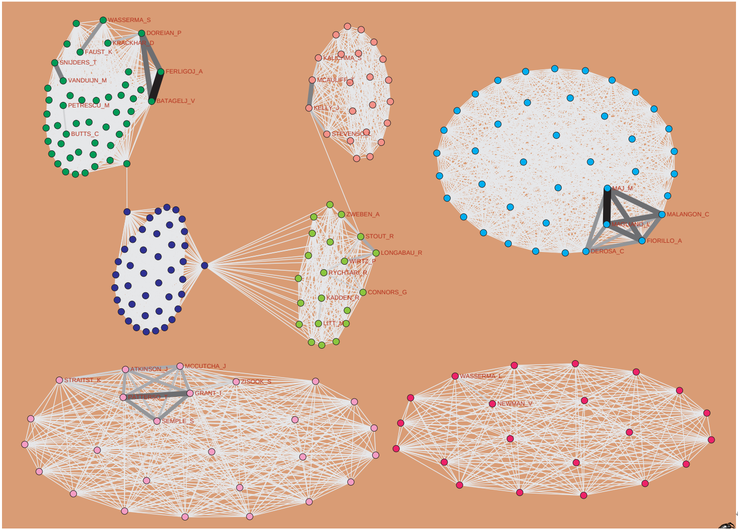
co_{ij} = the number of works that authors i and j wrote together
co_{ii} = the total number of works that author i wrote

It holds: $co_{ij} = co_{ji}$.

Problem - papers with many co-authors

Cores of orders 20–47 in **Co**(SN5)

Network SN5 (2008): for "social network*" + most frequent references + around 100 social networkers; $|W| = 193376$, $|C| = 7950$, $|A| = 75930$, $|J| = 14651$, $|K| = 29267$



Outer product decomposition

Let $x = [x_1, x_2, \dots, x_n]$ and $y = [y_1, y_2, \dots, y_m]$ be vectors. Their *outer product* $x \circ y$ is an $n \times m$ matrix defined as

$$x \circ y = [x_i \cdot y_j]_{n \times m}$$

Denoting $S_x = \sum_i x_i$ and $S_y = \sum_j y_j$ we get

$$S = \sum_{i,j} (x \circ y)_{ij} = \sum_i \sum_j x_i \cdot y_j = \sum_i x_i \cdot \sum_j y_j = S_x \cdot S_y$$

Therefore: $S_x = S_y = 1 \quad \Rightarrow \quad S = 1.$

It is easy to veryfy that the outer product decomposition holds

$$\mathbf{AK} = \mathbf{AW} * \mathbf{WK} = \sum_w \mathbf{WA}[w, \cdot] \circ \mathbf{WK}[w, \cdot], \qquad \mathbf{AW} = \mathbf{WA}^T$$

Note that for networks with all weights equal to 1 we have

$$\mathbf{WA}[w, \cdot] \circ \mathbf{WK}[w, \cdot] = K_{N_{WA}(w), N_{WK}(w)}$$

$x' = x / S_x \quad \Rightarrow \quad S_{x'} = 1$

Example

outer product decomposition

$$WA^T * WK = AK$$

WA	a1	a2	a3	a4
w1	1	0	1	0
w2	1	1	0	0
w3	1	0	1	1
w4	0	1	0	1
w5	1	0	1	1

T *

WK	k1	k2	k3	k4
w1	1	1	0	0
w2	1	0	1	0
w3	0	1	1	1
w4	0	0	1	0
w5	0	1	0	1

=

AK	k1	k2	k3	k4
a1	2	3	2	2
a2	1	0	2	0
a3	1	3	1	2
a4	0	2	2	2

AK	k1	k2	k3	k4
a1	2	3	2	2
a2	1	0	2	0
a3	1	3	1	2
a4	0	2	2	2

=

H1	k1	k2	k3	k4
a1	1	1	0	0
a2	0	0	0	0
a3	1	1	0	0
a4	0	0	0	0

+

H2	k1	k2	k3	k4
a1	1	0	1	0
a2	1	0	1	0
a3	0	0	0	0
a4	0	0	0	0

+

H3	k1	k2	k3	k4
a1	0	1	1	1
a2	0	0	0	0
a3	0	1	1	1
a4	0	1	1	1

+

H4	k1	k2	k3	k4
a1	0	0	0	0
a2	0	0	1	0
a3	0	0	0	0
a4	0	0	1	0

+

H5	k1	k2	k3	k4
a1	0	1	0	1
a2	0	0	0	0
a3	0	1	0	1
a4	0	1	0	1

Fractional approach

(Out) normalization $n(\mathcal{N})$ is a network obtained from $\mathcal{N}$ in which the weight of each arc a is divided by the sum of weights of all arcs having the same initial node as the arc a. For binary networks

$$n(\mathbf{A}) = \text{diag}(\frac{1}{\max(1, \text{outdeg}(i))})_{i \in \mathcal{I}} * \mathbf{A}$$

To get an equal contribution $S = 1$ of each work to the co-autorship network we have to use normalized vectors in the outer product decomposition. This is equivalent to define a normalized autorship network $\mathbf{N} = n(\mathbf{WA})$ – *fractional approach* [2, 4, 8, 6, 9].

Third co-authorship network

Then the *third co-authorship network* is

$$\mathbf{Ct} = \mathbf{N}^T * \mathbf{N}$$

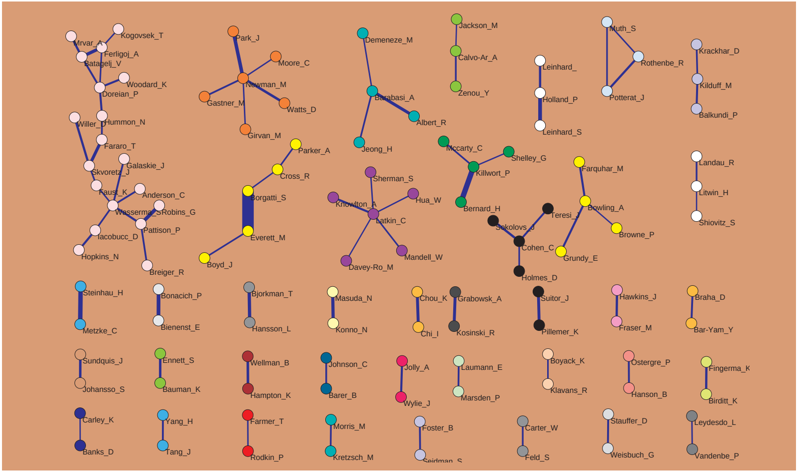
ct_{ij} = the total contribution of 'collaboration' of author i with author j to works.

It holds $ct_{ij} = ct_{ji}$.

We usually transform the network **Ct** into the corresponding undirected network with doubled weights.

Components in **Ct**(SN5) cut at level 0.5

Network SN5 (2008): for "social network*" + most frequent references + around 100 social networkers; $|W| = 193376$, $|C| = 7950$, $|A| = 75930$, $|J| = 14651$, $|K| = 29267$



Newman's co-authorship network

In 2001 Newman [7] proposed another fractional approach to defining a co-authorship network considered as a proxy for collaboration network. It is based on slightly different normalization

$$n'(\mathbf{A}) = \mathrm{diag}\left(\frac{1}{\max(1, \mathrm{outdeg}(i) - 1)}\right)_{i \in \mathcal{I}} * \mathbf{A}$$

The fourth or *Newman's co-authorship network* is defined as

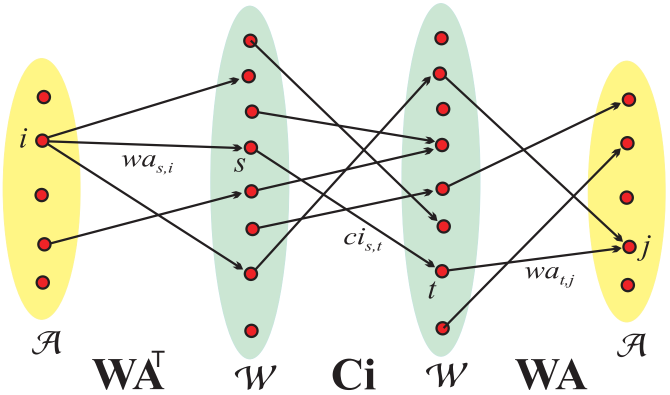
$$\mathbf{Ct}' = \mathbf{N}^T * \mathbf{N}', \text{ where } \mathbf{N}' = n'(\mathbf{WA}).$$

ct'_{ij} = the total contribution of 'strict collaboration' of authors i and j to works.

The final result is returned as an undirected simple network without loops and with weights

$$ct'_{ij} = \sum_p \frac{2 \cdot wa_{pi} \cdot wa_{pj}}{\max(1, \mathrm{outdeg}_{WA}(p)) \cdot \max(1, \mathrm{outdeg}_{WA}(p) - 1)}$$

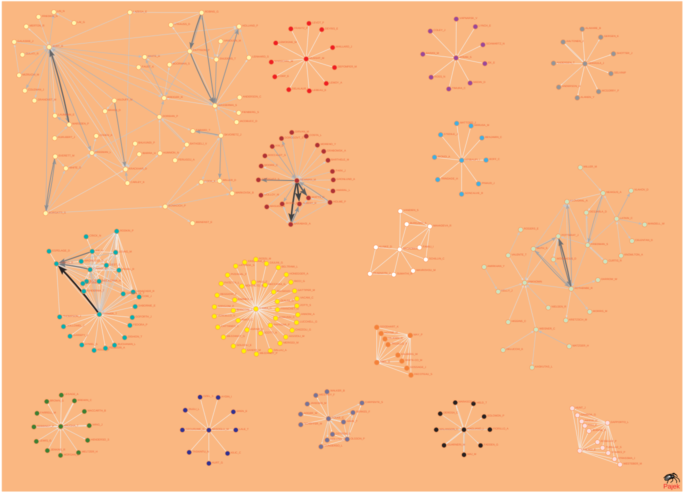
Authors' citations network



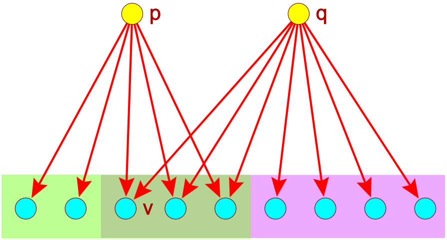
Ca $=$ **AW** $*$ **Ci** $*$ **WA** is a network of citations between authors. The weight $w(i, j)$ counts the number of times a work authored by i is citing a work authored by j.

Islands in SN5 authors citation network

Network SN5 (2008): for "social network*" + most frequent references + around 100 social networkers; $|W| = 193376$, $|C| = 7950$, $|A| = 75930$, $|J| = 14651$, $|K| = 29267$



Bibliographic Coupling



In WoS2Pajek the citation relation means p **Ci** q $\equiv$ work p cites work q.

Therefore the *bibliographic coupling* (Kessler, 1963) network **biCo** can be determined as

$$\mathbf{biCo} = \mathbf{Ci} * \mathbf{Ci}^T$$

$bico_{pq} = \#$ of works cited by both works p and q $= |\mathbf{Ci}(p) \cap \mathbf{Ci}(q)|$.

Bibliographic coupling weights are symmetric: $bico_{pq} = bico_{qp}$:

$$\mathbf{biCo}^T = (\mathbf{Ci} * \mathbf{Ci}^T)^T = \mathbf{Ci} * \mathbf{Ci}^T = \mathbf{biCo}$$

Bibliographic Coupling

fractional approach

Again we have problems with works with many citations, especially with review papers. To neutralize their impact we can introduce normalized measures. Let's first look at

$$\mathbf{biC} = n(\mathbf{Ci}) * \mathbf{Ci}^T$$

where $n(\mathbf{Ci}) = \mathbf{D} * \mathbf{Ci}$ and $\mathbf{D} = \text{diag}(\frac{1}{\max(1,\text{outdeg}(p))})$. $\mathbf{D}^T = \mathbf{D}$.

$$\mathbf{biC} = (\mathbf{D} * \mathbf{Ci}) * \mathbf{Ci}^T = \mathbf{D} * \mathbf{biCo}$$

$$\mathbf{biC}^T = (\mathbf{D} * \mathbf{biCo})^T = \mathbf{biCo}^T * \mathbf{D}^T = \mathbf{biCo} * \mathbf{D}$$

For $\mathbf{Ci}(p) \neq \emptyset$ and $\mathbf{Ci}(q) \neq \emptyset$ it holds (proportions)

$$\mathbf{biC}_{pq} = \frac{|\mathbf{Ci}(p) \cap \mathbf{Ci}(q)|}{|\mathbf{Ci}(p)|} \quad \text{and} \quad \mathbf{biC}_{qp} = \frac{|\mathbf{Ci}(p) \cap \mathbf{Ci}(q)|}{|\mathbf{Ci}(q)|} = \mathbf{biC}^T_{pq}$$

and $\mathbf{biC}_{pq} \in [0, 1]$.

Bibliographic Coupling

fractional approach

Using **biC** we can construct different normalized measures such as

$$\mathbf{biCog}_{pq} = \sqrt{\mathbf{biC}_{pq} \cdot \mathbf{biC}_{qp}} = \frac{|\mathbf{Ci}(p) \cap \mathbf{Ci}(q)|}{\sqrt{|\mathbf{Ci}(p)| \cdot |\mathbf{Ci}(q)|}}$$

Geometric mean
Salton cosinus

$$\mathbf{biCoj}_{pq} = (\mathbf{biC}_{pq}^{-1} + \mathbf{biC}_{qp}^{-1} - 1)^{-1} = \frac{|\mathbf{Ci}(p) \cap \mathbf{Ci}(q)|}{|\mathbf{Ci}(p) \cup \mathbf{Ci}(q)|}$$

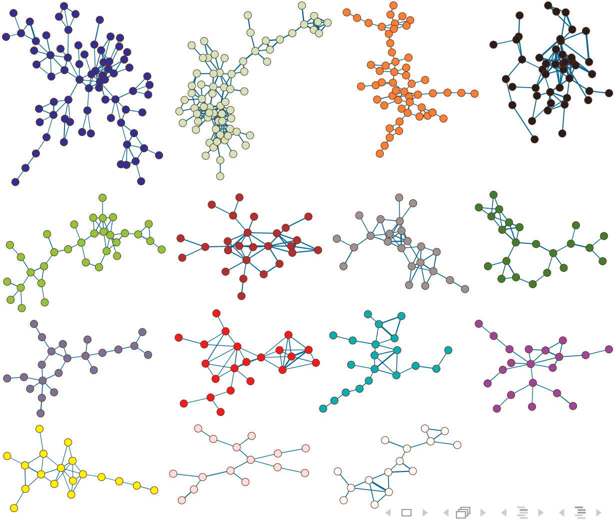
Jaccard index

Both measures are symmetric.

Bibliographic Coupling

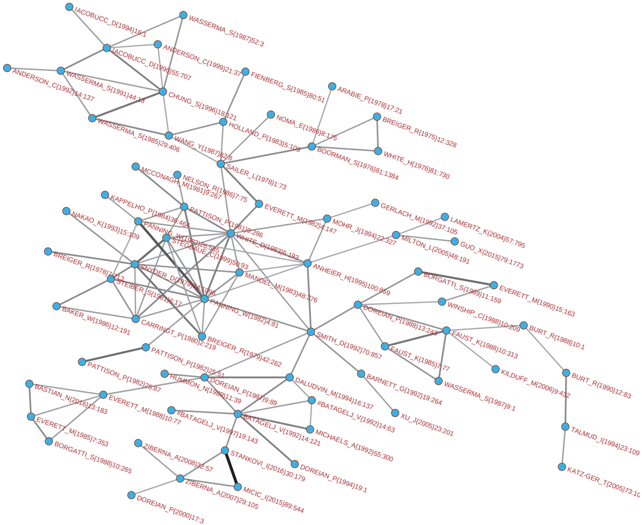
Jaccard islands [15, 75]

Network `BMc` (2016): for `"block model*"` or `"network cluster*"` `...`;
$|W| = 5695$, $|A| = 13376$, $|J| = 1756$, $|K| = 10269$



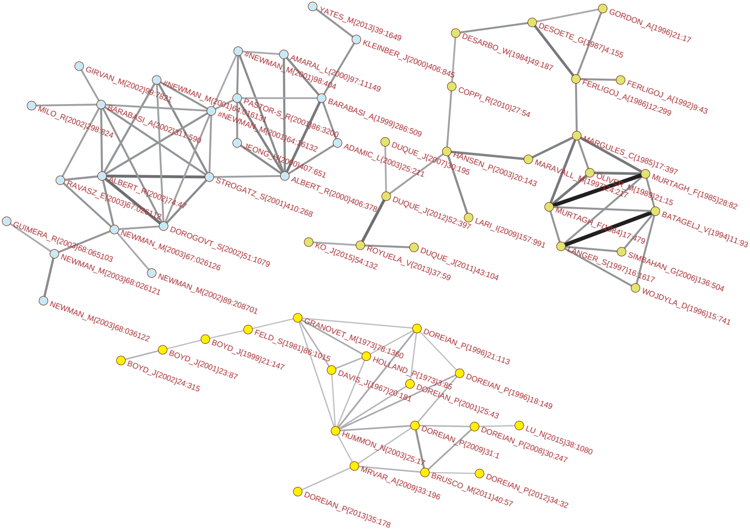
Bibliographic Coupling

Jaccard island 4 (74)



Bibliographic Coupling

Jaccard islands 12 (23), 11 (22), 1 (18)



Conclusions

- Network multiplication enables us to link by derived networks some directly unlinked modes in a multimode network.
- The analysis of the obtained networks can be based on their weights using cuts, (generalized) cores, islands, etc.
- It is important to understand the meaning of the weights. Weights appropriate for our research question can be often obtained by an appropriate normalization.

References I

Batagelj, V., Doreian, P., Ferligoj, A., Kejžar, N. (2014). Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution. Wiley Series in Computational and Quantitative Social Science. Wiley.

Batagelj, V, Cerinšek, M (2013): On bibliographic networks. Scientometrics 96(3), 845-864. paper

Carley, K.M. (2003). Dynamic Network Analysis. in the Summary of the NRC workshop on Social Network Modeling and Analysis, Breiger, R. and Carley, K.M. (Eds.), National Research Council. preprint

Cerinšek, M., Batagelj, V. (2015): Network analysis of Zentralblatt MATH data. Scientometrics, 102(1), 977-1001. paper

Krackhardt, D., Carley, K.M. (1998). A PCANS Model of Structure in Organization. In Proceedings of the 1998 International Symposium on Command and Control Research and Technology Evidence Based Research: 113-119, Vienna, VA. MetaMatrix, paper

References II

- Leydesdorff, L. and Park, H.W. (2017). Full and Fractional Counting in Bibliometric Networks. Journal of Informetrics 11(1), 117–120. arXiv, paper
- Newman, M.E.J. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality Physical Review E, 64, 016132.
- Perianes-Rodriguez, A., Waltman, L., Van Eck, N.J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. Journal of Informetrics, 10(4), 1178-1195. paper
- Prathap, G. and Mukherjee, S. (2016). A conservation rule for constructing bibliometric network matrices. arXiv
- Žiberna, A. (2016). Blockmodeling linked networks. Abstracts for International Conference Applied Statistics, Ribno, Slovenia, p. 45.