Analysis of enriched co-authorship networks: methodology and a case study

Miloš Savić

University of Novi Sad, Serbia
Faculty of Sciences,
Department of Mathematics and Informatics

Outline

- Introduction
- Methodology
- Case study
- Main results
- Conclusions

Outline

- **Introduction**
- Methodology
- Case study
- Main results
- Conclusions

Research collaboration

- **Collaboration:** key social feature of modern science
- **Science from a social perspective:** complex self-organizing social system
- **Katz:** "scientific collaboration is a social process and probably there are as many reasons for researchers to collaborate as there are reasons for people to communicate"
  - growth of scientific community
  - professionalization of research practice
  - rationalization of scientific manpower
  - information & knowledge overload
  - complexity of contemporary research problems
  - cross-fertilization of ideas, etc.

Research collaboration

- **Research collaboration can be studied at various levels:**
  - intra-institutional (within an institution)
  - inter-institutional (between institutions)
  - national (within a country)
  - international (between countries)
  - disciplinary (within a scientific discipline)
  - inter-disciplinary (between scientific disciplines)
- **Major research questions:**
  - how research collaboration is structured?
  - how the structure of research collaboration evolves?
  - how research collaboration is related to research productivity and impact of multi-authored publications?

Research collaboration

- Research collaboration may manifest in various formal and informal forms
- Co-authorship — the most visible and well-documented manifestation of scientific collaboration
  - availability of massive bibliographic databases
- Co-authorship networks — social networks encompassing researchers
  - Nodes — researchers
  - A and B are connected if A and B co-authored at least one publication (with or without other co-authors)
  - Link weights — the strength of research collaboration

Link weighting schemes

- Straight scheme
  - w(x, y) = the number of joint publications of x and y
- Salton's scheme — a normalized variant of the straight scheme

$$w = \frac{h_{x,y}}{\sqrt{h_x \cdot h_y}}$$

h(x) — the number of publications (co-)authored by x
h(y) — the number of publications (co-)authored by y
h(x, y) — the number of joint publications of x and y

- Newman's scheme
  - More authors a paper has less weight should be added to the total strength of research collaboration

$$w = \sum_{k \in J} \frac{1}{n_k - 1}$$

J — the set of joint publications of x and y
n(k) — the number of authors of publication k

Co-authorship networks as directed graphs

- **Yoshikane et al., 2006:**
  - w(x → y) = the number of joint publications of x and y in which y is the first author
  - leading and following roles in research collaboration
  - application of weighted Kleinberg's HITS centrality measures to identify research authorities and important followers

$$H(x) = \sum_{y \in O(x)} \left(A(y) \times \text{weight}(x \rightarrow y) \right)$$

$$A(x) = \sum_{z \in I(x)} \left(H(z) \times \text{weight}(z \rightarrow x) \right)$$

Co-authorship networks as directed graphs

- **Liu et al., 2005:** two co-authors connected by a pair of reciprocal directed links
  - exclusivity of x and y on joint article p $e(x,y,p) = 1/(n_p - 1)$
  - co-authorship frequency of x with y

$$c(x,y) = \sum_{p \in P_x} e(x,y,p) \qquad w(x \to y) = \frac{c(x,y)}{\sum_{k \in V \setminus \{x\}} c(x,k)}$$

  - **AuthorRank:** a weighted variant of PageRank to identify important authors

$$\mathrm{AR}(x) = (1-d) + d \sum_{z \in I_x} \left(\mathrm{AR}(z) \times \mathrm{weight}(z \to x)\right)$$

Applications of co-authorship networks

- **Main application: to study the structure and evolution of research collaboration**
  - Centrality metrics — identification of central, the most important researchers in a research community
  - Community detection algorithms — identification and analysis of research groups
  - Statistical properties of node degree/strength distributions — inequalities and gaps in research collaboration
  - Characteristics of connected components and subgraphs — cohesiveness of research collaboration
  - Trends and patterns in the evolution of scientific collaboration at the level of researchers, research groups and scientific fields
- **Other applications:** prediction and recommendation of research collaboration, recommendation of reviewers for a manuscript, prediction of scientific impact of researchers

Analysis of co-authorship networks

- Mark Newman (2001): a general methodological framework to analyze large-scale co-authorship networks
  - Metrics and methods of complex network theory

- Frequently observed characteristics
  - Heavy-tailed distributions of node centrality and productivity metrics
  - Giant connected components
  - Assortative mixing patterns
  - Small-worlds emphasized by shrinking diameters
  - The funneling effect
  - High degree of local clustering and community structures
  - Evolution governed by the preferential attachment principle
  - Strong correlations between centrality metrics and citation counts

Extraction of co-authorship networks

- Trivial for bibliographic databases in which authors are uniquely identified

- **Name disambiguation problem** if authors are identified by their names in bibliographic records
  - it is hard to identify nodes of co-authorship network
  - **Name homonymy** — different individuals having the same name
  - **Name synonymy** — a single individual with multiple name variants in bibliographic records
    - spelling errors, orthographic variants, abbreviated names, transliteration, pen names, marriage, etc.

Approaches to the name disambiguation problem

- **Newman's initials based approaches**
  - authors identified by full last name, first name reduced to the first/all initials
  - used in seminal Newman's papers on co-authorship networks
  - handle only name synonymy when spelling errors and other inconsistencies appear in first names, increase the level of name homonymy
- **Heuristic approaches**
  - Name synonymy: string similarity metrics and node distance in co-authorship network
  - Name homonymy: articulation points in ego-networks

Machine learning approaches to ND

- S = a set of references of authors with identical or highly similar name labels
- **Author assignment methods**
  - S’ = S annotated with true identities of authors
  - a classification model trained on S’
  - only identities appearing in S’ can be assigned to references in which true identities are unknown
- **Author grouping methods**
  - cluster S, distinct persons determined by clusters in S (one cluster — one person)
  - reference similarity functions: predefined (e.g. string similarity) or learned (classifiers trained on a dataset containing co-referent and non-co-referent reference pairs)

Our contributions

- Analysis of **enriched intra-institutional co-authorship networks**
  - Networks representing research collaboration within an institution
  - Underlying institutional structure which defines one partitioning of the network
  - Nodes enriched with metrics quantifying different determinants of research performance and demographic attributes (e.g. gender)

- Case study: FS-UNS co-authorship network
  - Extracted from bibliographic records contained in the institutional CRIS-UNS system
  - Benefits of our CRIS system
    - No name disambiguation problems
    - Categorization of publications by the rule book prescribed by the Serbian Ministry of Science
      → Serbian research competency index metric

Outline

- Introduction
- **Methodology**
- Case study
- Main results
- Conclusions

- **Hybrid methodology to analyze intra-institutional research collaboration**
  - Domain-independent metrics and methods of complex network theory
  - Domain-dependent metrics of researcher productivity
  - Non-parametric statistical tests applied to the set of metric values of independent groups of nodes
  - Graph representations derived from enriched co-authorship networks

- Basic steps
  1. Connected component analysis to evaluate cohesiveness of the institution
  2. Analysis of cohesiveness of research departments
  3. Analysis of inter-department links and researchers involved in inter-department collaborations
  4. Identification of research groups using community detection techniques, analysis of inter-group collaborations
  5. Comparison of departments and research groups relying on researcher evaluation metrics

Metric-based comparison of node groups

- Non-parametric tests of stochastic superiority
  - Mann-Whitney U (MWU)
  - Kruskal-Wallis ANOVA with MWU as post-hoc test

- Probabilities of superiority
  - M — an arbitrary node metrics quantifying some aspect of research performance
  - M(G1) and M(G2) — the sets of M values for two independent groups of nodes G1 and G2
  - g1 — randomly selected value from M(G1)
  - g2 — randomly selected value from M(G2)
  - **P(g1 > g2) — the probability of superiority of G1 over G2 regarding the aspect of research performance quantified by M**
  - P(g2 > g1) — the opposite probability of superiority
  - P(g1 > g2) + P(g2 > g1) may be < 1, i.e. P(g1 = g2) may be > 0

Metric-based comparison of node groups

- Non-parametric tests of stochastic superiority used to examine differences between:
  - Researchers involved in intra-department collaborations and researchers who collaborate only with colleagues from their own departments
  - Researchers belonging to different departments
  - Researchers from the core of the network and researchers from the periphery (if the network has a core-periphery structure)
  - Researchers belonging to different research groups
  - Researchers whose collaboration is bounded to their own research group and researchers involved in inter-group research collaboration
  - Male and female researchers

Identification of core and periphery researchers

- k-core decomposition
- k-core: maximal subgraph K such that the degree of every node is >= k in K
- nested core-periphery structure: k-cores are connected graphs or have giant connected components
- shell-index(x) = k if x belongs to k-core, but not to (k+1)-core
- **core researchers — the minimal subset C of nodes such that the sum of shell indexes of nodes in C is higher than the sum of shell indexes of the rest of nodes**
- Efficient k-core decomposition by the Batagelj-Zaveršnik algorithm

Algorithm 2.1: The Batagelj-Zaveršnik algorithm for k-core decomposition

```
input  : a graph G = (V, E)
output: S – an array of integers such that S[v] (v ∈ V) is equal to the shell index of v

m = the maximal degree of nodes in V
d = an array of integers of length n where n = |V|
D = an array of m empty node sets

foreach v ∈ V do
    k = the degree of node v
    d[v] = k
    add v to D[k]
end

for k = 0 to m do
    while D[k] ≠ ∅ do
        x = remove a random node from D[k]
        S[x] = k
        foreach v ∈ V : {x, v} ∈ E do
            if d[v] > k then
                remove v from D[d[v]]
                add v to D[d[v] − 1]
                d[v] = d[v] − 1
            end
        end
    end
end
```

Inter-department/group collaborations

- Analysis of the structure of block models determined by partitions of the network according to
  - departamental affiliation of researchers
  - cohesive node groups identified by a community detection algorithm
    - Various community detection techniques applied
    - The best partitioning selected using
      - Q — the weighted modularity measure
      - r — the ratio between the total weight of inter-community and intra-community links
      - Q(A) > Q(B) and r(A) < r(B) → partition A better than B
    - **Strong links:** the minimal subset C such that the total weight of C is higher than the total weight of the rest of links

- **GMO — greedy modularity optimization**
  - agglomerative clustering, two communities whose join maximally increase (minimally decrease) modularity are merged
- **LV — Louvain**
  - agglomerative clustering, modularity optimized locally (by moving nodes to neighbor communities) then on the block model
- **WT — Walktrap**
  - agglomerative clustering, distance(X, Y) = the probability that a random walk starting at X reaches Y in k steps
- **EB — Girvan-Newman edge-betweenness algorithm**
  - divisive clustering by iteratively removing links having the highest betweenness centrality
- **IM — Infomap**
  - reveals communities by optimally compressing random walks
- **SOM — spectral modularity optimization**
  - divisive clustering according to the leading eigenvector of the modularity matrix Q[i, j] = A[i, j] - E(A[i, j])

Group superiority graph (GSG)

- M — an arbitrary researcher evaluation metric
- GSG corresponding to M
  - directed graph
  - nodes: research groups
  - A → B: A strongly dominates over B w.r.t M
    - MWU(M(A), M(B)) rejected
    - PS(A > B) > 0.75
  - isolated nodes removed
  - Inferior groups: out-degree = 0 in GSG
  - Superior groups: in-degree = 0 in GSG
  - Links are transitive: A → B and B → C imply A → C
  - Empty GSG — no strong differences between research groups w.r.t M

Cohesiveness of research collaboration

- For the whole institution
  - Identification and analysis of connected components using metrics from complex network theory
  - Existence of a giant connected component?
- For research departments and research groups
  - Are departments/groups good clusters in the network?
  - Graph clustering evaluation metrics: internal density, weighted conductance (WC) and weighted Flake degree fraction (WFDF)
  - The Radicchi notion of communities in complex networks

$$\text{D is Radicchi weak cluster} \iff \sum_{r \in D} w^{\text{intra}}(r) > \sum_{r \in D} w^{\text{inter}}(r)$$

$$\iff \text{WC}(D) < 0.5$$

$$\text{D is Radicchi strong cluster} \iff (\forall r \in D)\, w^{\text{intra}}(r) > w^{\text{inter}}(r)$$

$$\iff \text{WFDF}(D) = 1.0$$

Outline

- Introduction
- Methodology
- **Case study**
- Main results
- Conclusions

Case study

- FS-UNS co-authorship network
  - Extracted from the CRIS UNS data
    - 14986 publications by 423 FS-UNS researchers and 5690 external co-authors
  - The Newman schema used to assign link weights

Department	Abbrv.
Department of Biology and Ecology	DBE
Department of Physics	DP
Department of Geography, Tourism and Hotel Management	DG
Department of Chemistry, Biochemistry and Environmental Protection	DC
Department of Mathematics and Informatics	DMI

Department	Researchers	Male [%]	Female [%]	Avg. age
DMI	87	49.43	50.57	45.3
DG	66	57.58	42.42	42.9
DBE	118	25.42	74.58	41.2
DP	57	56.14	43.86	46.5
DC	95	24.21	75.79	42.7

Metric	Abbreviation	Category
Productivity, normal count	PRON	Productivity
Productivity, fractional count	PROF	Productivity
Productivity, straight count	PROS	Productivity
Serbian Research Competency Index	SRCI	Productivity
The total number of co-authors	COLL	Collaboration
The number of FS-UNS co-authors	LCOLL	Collaboration
The number of external co-authors	ECOLL	Collaboration
The strength of research collaboration with all co-authors	WCOLL	Collaboration
The strength of research collaboration with FS-UNS co-authors	WLCOLL	Collaboration
The strength of research collaboration with external co-authors	WECOLL	Collaboration
Clustering coefficient	CC	Collaboration
The degree of intra-group collaboration	IntraDEG	Collaboration
The degree of inter-group collaboration	InterDEG	Collaboration
The strength of intra-group collaboration	WIntraDEG	Collaboration
The strength of inter-group collaboration	WInterDEG	Collaboration
Betweenness centrality	BET	Importance
Weighted betweenness centrality	WBET	Importance
Closeness centrality	CLO	Importance
Weighted closeness centrality	WCLO	Importance
Eigenvector centrality	EVC	Importance

SRCI (Serbian Research Competency Index)

- Categories of publication venues prescribed by the Serbian ministry of Science
  - more than 20 categories, e.g.
    - article at an international conference — 1 point,
    - article in the top 10% SCI ranked journals — 10 points
  - separate categories for international and domestic journals, as well as for international and domestic conferences
- SRCI(X) - the sum of points publications (co-)authored by X
- SRCI does not take into account the number of authors of a publication, each of them received the full credit

Outline

- Introduction
- Methodology
- Case study
- **Main results**
- Conclusions

Connected component analysis

- 423 researchers in 15 connected components,
- 14 connected components are isolated nodes
  - FS-UNS researchers whose entire production consists of solo-authored articles
- Giant connected component encompassing 96.7% nodes → FS-UNS researchers form a cohesive institutional research community
- The small-world property in the Watts-Strogatz Sense
  - SW ≈ SW(rnd), CC >> CC(rnd)

Cohesiveness of research departments

- All FS-UNS departments are Radicchi weak and close to Radicchi strong clusters in the network
  - Intra-department collaborations are stronger than inter-department collaborations for a large majority of researchers but not for all of them
- The strongest intra-department collaborations: DP and DC
- The weakest intra-department collaborations: DMI
- The most closed department: DG (highest internal density, lowest conductance)

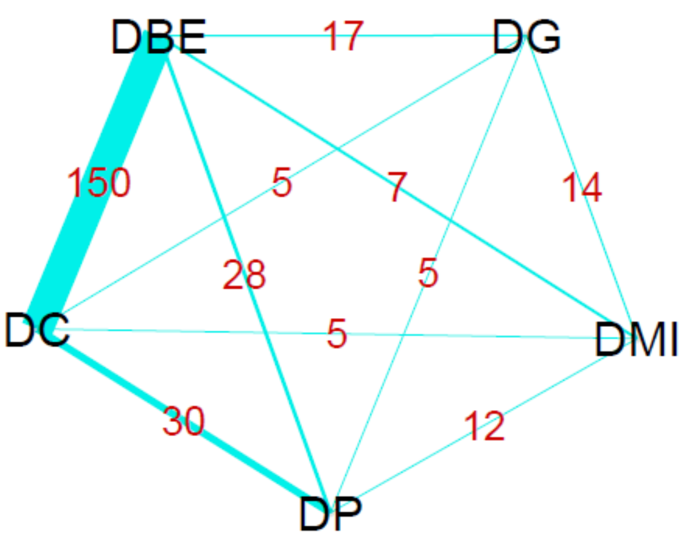
Metric	DBE	DP	DC	DMI	DG
The number of researchers	118	57	95	87	66
The number of non-trivial components	1	1	1	1	1
The number of isolated nodes	3	3	0	7	6
The number of intra-department links	660	240	617	197	560
The number of inter-department links	412	174	411	71	96
Internal density	0.096	0.15	0.14	0.05	0.26
Total weight of intra-department links	8073	5636	9261	1532	2513
Total weight of inter-department links	1607	683	1825	195	160
Average internal degree	11.19	8.42	12.99	4.53	16.97
Average internal weighted degree	136.83	197.76	194.97	35.22	76.15
Weighted conductance	0.17	0.11	0.16	0.11	0.06
Weighted Flake degree fraction	0.97	0.93	0.98	0.95	0.95

Inter-department collaborations

- Researchers involved in inter-department collaborations are drastically more productive, collaborative and institutionally important

Node metric	Avg(G_1)	Avg(G_2)	U	p	PS_1	PS_2
SRCI	160.378	58.6939	11178.5	1.08E-18*	0.7482	0.2507
PRON	104.9031	32.9031	10333	2.06E-21*	0.764	0.2285
PROS	29.2555	13	13781	1.40E-11*	0.6764	0.2959
PROF	27.9682	12.3087	13477.5	2.69E-12*	0.697	0.3029
LCOLL	18.7225	7.4592	7486.5	4.92E-32*	0.82	0.1566
ECOLL	51.0088	13.4745	8411.5	2.52E-28*	0.8038	0.1819
COLL	69.7313	20.9337	7360	1.62E-32*	0.8304	0.1612
BET	769.6687	98.0929	7775	5.09E-31*	0.8166	0.1661

- The departmental collaboration network of FS-UNS is a clique, but the strengths of inter-department collaborations are highly unbalanced (a lot of space to improve inter-department collaborations)

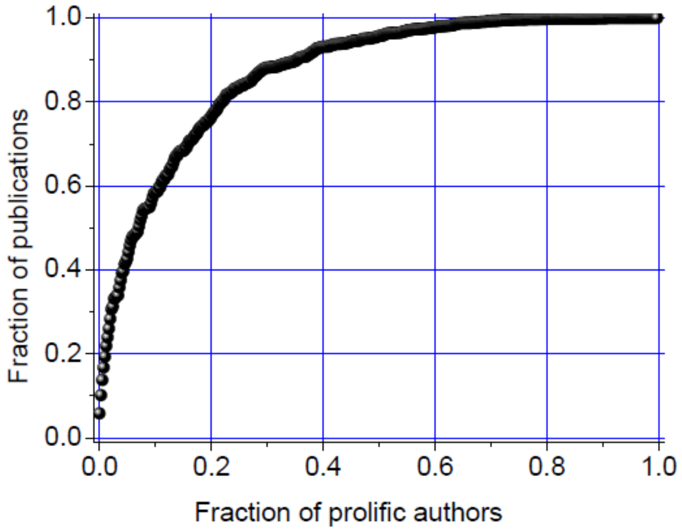


Researcher metrics

- The distributions of researcher metrics are highly skewed to the right
  → deep inequalities in productivity, collaboration and inst. importance

Metric	Mean	Standard deviation	Skewness	Max.
SRCI	113.26	142.7	2.76	1124.70
PRON	71.54	95.9	3.27	871
PROS	21.72	29.9	2.92	242
PROF	20.71	28.76	3.4	266
LCOLL	13.50	10.91	1.11	56
ECOLL	33.62	40.94	2.23	260
COLL	47.12	49.17	1.98	311
BET	458.49	913.17	3.98	9064.98

- The Pareto principle: 80% of publications authored by 20% of researchers



Metric-based comparison of departments

- Statistically significant differences (SSD) present regarding SRCI and PRON, but absent regarding PROS and PROF
  - **SRCI and PRON – biased measures of productivity**
- SSD in both local and external collaboration
- No SSD regarding institutional importance

Metric	DBE	DP	DC	DMI	DG	χ^2	p-value
SRCI	91.86	174.58	151.84	94.96	67.17	26.01	3.15E-05*
PRON	74.77	98.37	90.54	44.75	50.58	22.68	1.47E-04*
PROS	19.74	25.68	23.48	21.3	19.88	7.85	0.097
PROF	19.17	23.48	20.69	22.2	19.15	6.38	0.172
LCOLL	14.68	11.47	17.32	5.34	18.42	99.11	1.52E-20*
ECOLL	39.17	41.65	43.59	12.36	30.42	49.11	5.54E-10*
COLL	53.85	53.12	60.91	17.7	48.85	69.71	2.61E-14*
BET	514.21	464.53	362.39	553.4	366.87	3.24	0.51811

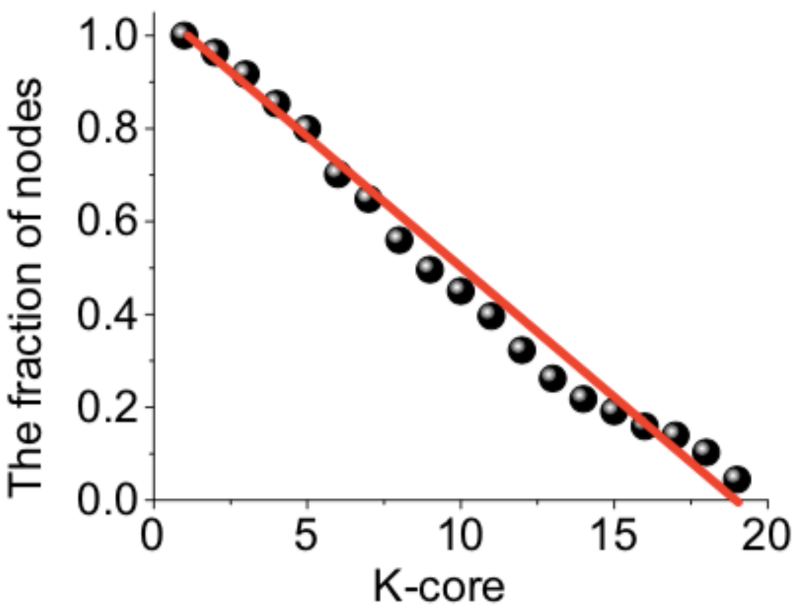
Post-hoc pairwise comparison

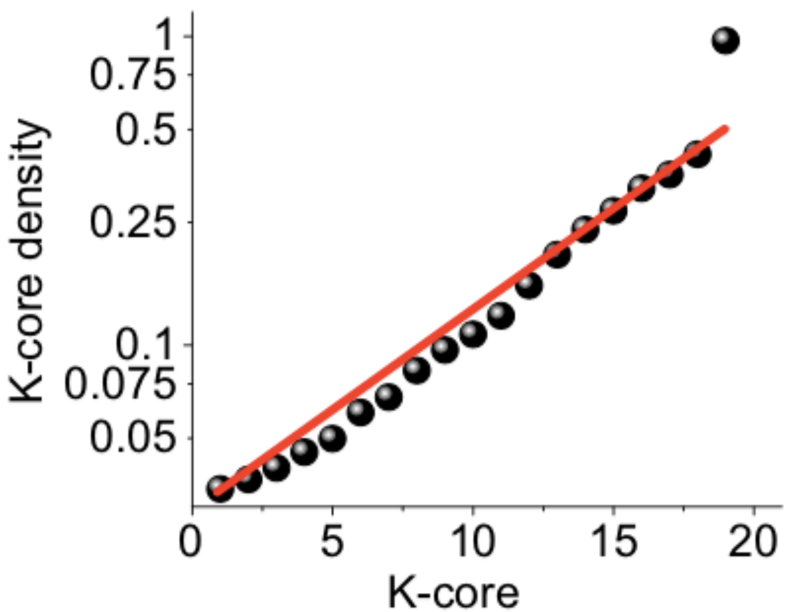
- DP and DC: superior regarding SRCI and PRON
- DMI: the lowest degree of both local and external research collaboration
- DC and DG: active stimulation of intra-institutional collaboration

Metric	Department 1	Department 2	U	p-value	PS_1	PS_2
SRCI	DG	DP	1280	0.0023	0.34	0.66
	DBE	DP	2516.5	0.0071	0.37	0.62
	DP	DMI	1825.5	0.0076	0.63	0.37
	DG	DC	1956	0.0001	0.31	0.69
	DBE	DC	4046.5	0.0005	0.36	0.64
	DMI	DC	2884	0.0004	0.35	0.65
PRON	DP	DMI	1824.5	0.0075	0.63	0.36
	DG	DC	2236.5	0.002	0.35	0.64
	DBE	DC	4345	0.0048	0.38	0.61
	DMI	DC	2481	$< 10^{-4}$	0.29	0.69
LCOLL	DG	DBE	2913	0.0046	0.62	0.36
	DG	DP	1094.5	0.0001	0.70	0.28
	DG	DMI	851	$< 10^{-4}$	0.84	0.14
	DBE	DMI	2297.5	$< 10^{-4}$	0.75	0.20
	DP	DMI	1150.5	$< 10^{-4}$	0.75	0.21
	DBE	DC	4521	0.0153	0.39	0.58
	DP	DC	1888.5	0.0018	0.33	0.63
	DMI	DC	1073	$< 10^{-4}$	0.12	0.86
ECOLL	DG	DMI	1576	$< 10^{-4}$	0.71	0.26
	DBE	DMI	2906	$< 10^{-4}$	0.70	0.27
	DP	DMI	1319	$< 10^{-4}$	0.72	0.25
	DMI	DC	1879	$< 10^{-4}$	0.22	0.76

k-core decomposition results

- The FS-UNS co-authorship network has a strong and balanced nested core-periphery structure
  - 19 cores, all of them being connected subgraphs in the network
  - the density of cores increases exponentially
  - the fraction of nodes in k-cores decreases linearly with k
  - **Core researchers: shell-index >= 12 (32% of the total number)**





Core VS Peripheral Researchers

- Core researchers are drastically more productive, collaborative and institutionally important than peripheral researchers.
- Core researchers have more significant brokerage role within their ego-networks

Metric	Avg(C)	Avg(P)	U	p	NHA	PS_1	PS_2
PRON	124.2576	49.6354	7954	$< 10^{-4}$	no	0.78	0.22
PROF	31.7894	16.1697	10003	$< 10^{-4}$	no	0.73	0.27
PROS	32.4924	17.3430	10647	$< 10^{-4}$	no	0.70	0.28
SRCI	172.8083	89.7819	9684	$< 10^{-4}$	no	0.74	0.26
COLL	88.2273	29.8051	4653	$< 10^{-4}$	no	0.87	0.13
LCOLL	26.4697	8.0072	784.5	$< 10^{-4}$	no	0.98	0.02
ECOLL	61.7576	21.7978	7191	$< 10^{-4}$	no	0.80	0.19
WCOLL	120.2045	46.2960	7723	$< 10^{-4}$	no	0.79	0.21
WLCOLL	66.9061	22.8109	6641	$< 10^{-4}$	no	0.82	0.18
WECOLL	53.2984	23.4851	10089.5	$< 10^{-4}$	no	0.72	0.28
BET	813.9461	312.2748	8040	$< 10^{-4}$	no	0.78	0.22
CLO	0.3457	0.2897	4622	$< 10^{-4}$	no	0.87	0.13
EVC	0.0046	0.0014	849	$< 10^{-4}$	no	0.98	0.02
WBET	563.0291	221.8833	12644.5	$< 10^{-4}$	no	0.54	0.23
WCLO	0.6649	0.5056	7099	$< 10^{-4}$	no	0.81	0.19
CC	0.4659	0.5829	13654	$< 10^{-4}$	no	0.37	0.63

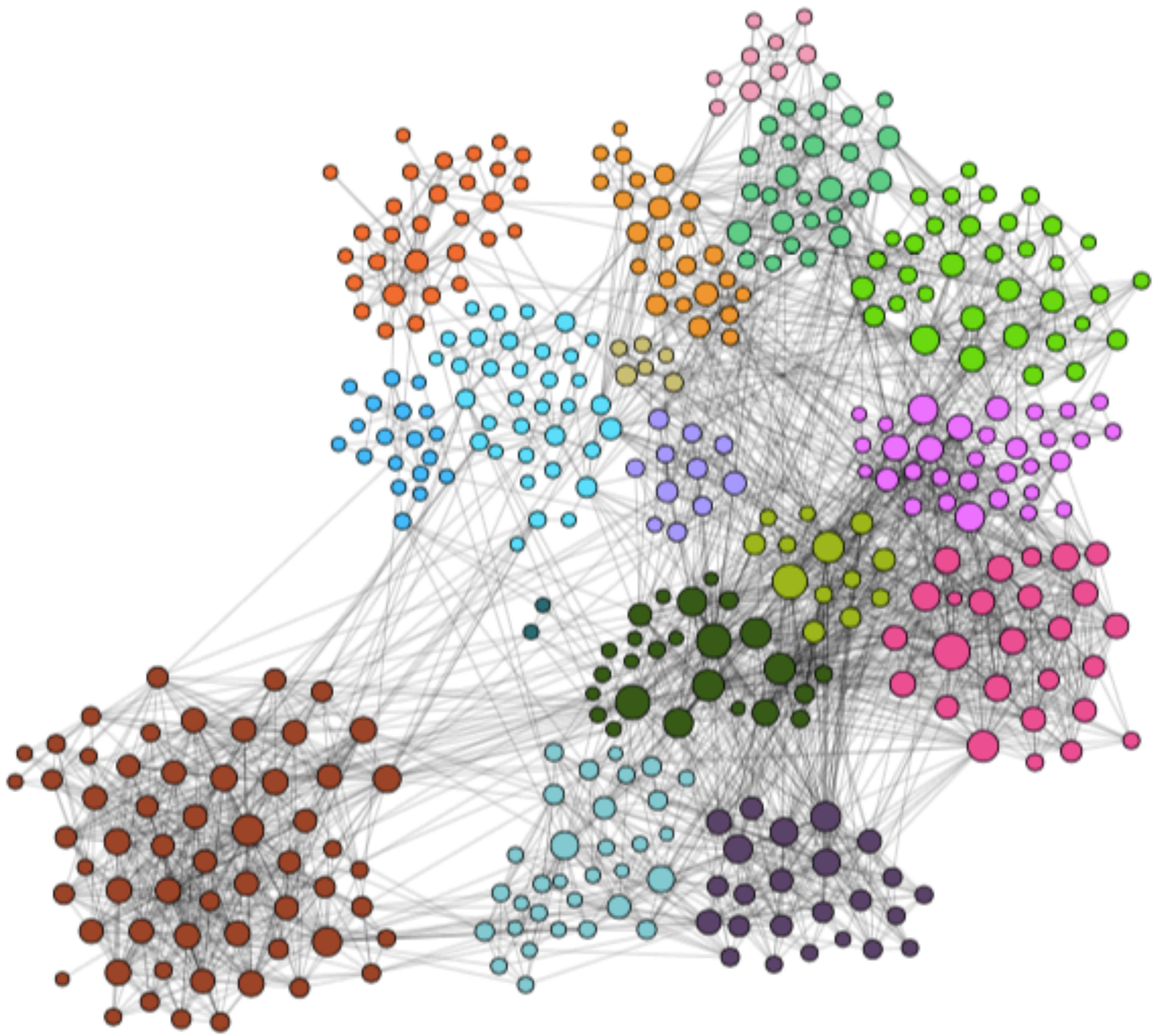
Identification or research groups

Algorithm	Reference	Q	NC	w^{intra}	w^{inter}	r
GMO	[5]	0.8371	18	6919.45	655.66	0.0947
IM	[21]	0.8141	41	6618.53	956.58	0.1445
LV	[3]	0.8466	17	6920.37	654.74	0.0946
WT	[18]	0.8207	37	6873.07	702.04	0.1021
EB	[8]	0.5486	13	5248.49	2326.63	0.4433
SOM	[16]	0.6022	27	6466.84	1108.28	0.1714

- the best performing algorithm: LV (Louvain)
  - the highest modularity, the lowest ratio of w(inter) and w(intra)
- agglomerative clustering techniques better than divisive

$$\text{LV} \succ \text{GMO} \succ \text{WT} \succ \text{IM} \succ \text{SOM} \succ \text{EB}$$

Fig. 8.3: The visualization of the FS-UNS co-authorship network after community detection by the Louvain algorithm. Nodes in the same color belong to the same community. The size of a node is proportional to its degree centrality.

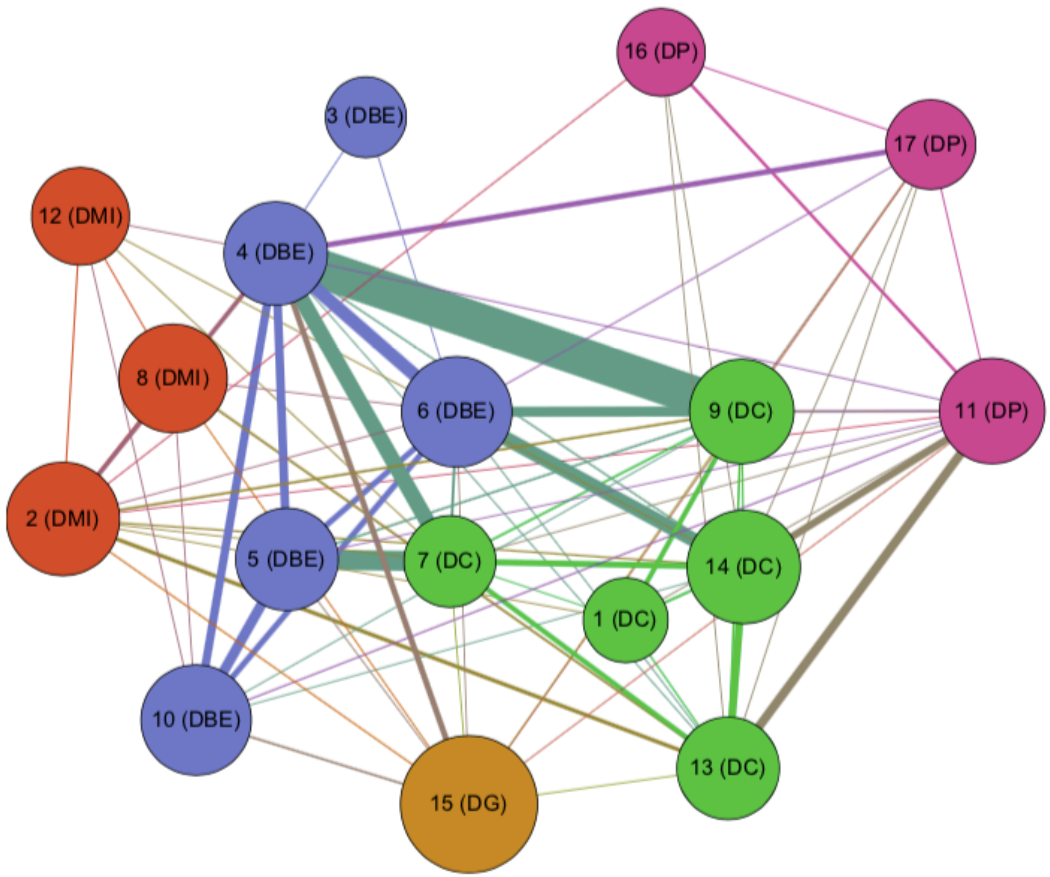


Research groups identified by Louvain

ID	Size	DBE	DP	DC	DMI	DG	DD	w^{intra}	w^{inter}	RS
C1	6	0	0	6	0	0	DC	179.33	23.58	no
C2	35	0	1	1	32	1	DMI	670.30	48.75	no
C3	2	2	0	0	0	0	DBE	82.57	2.70	yes
C4	25	23	0	1	0	1	DBE	927.73	260.12	no
C5	24	24	0	0	0	0	DBE	1230.69	117.35	no
C6	32	32	0	0	0	0	DBE	797.62	136.49	no
C7	13	0	0	13	0	0	DC	712.20	131.51	yes
C8	30	0	1	0	29	0	DMI	843.72	16.83	yes
C9	26	1	0	25	0	0	DC	1848.67	163.32	yes
C10	32	32	0	0	0	0	DBE	927.99	73.69	yes
C11	27	0	21	5	0	1	DP	761.24	68.34	yes
C12	19	0	0	0	19	0	DMI	299.17	10.23	yes
C13	24	0	10	13	1	0	DC	537.04	82.45	yes
C14	35	2	2	31	0	0	DC	948.64	107.32	yes
C15	59	0	0	0	0	59	DG	1642.18	30.58	yes
C16	9	0	9	0	0	0	DP	321.13	11.88	yes
C17	11	0	11	0	0	0	DP	1110.53	24.35	yes

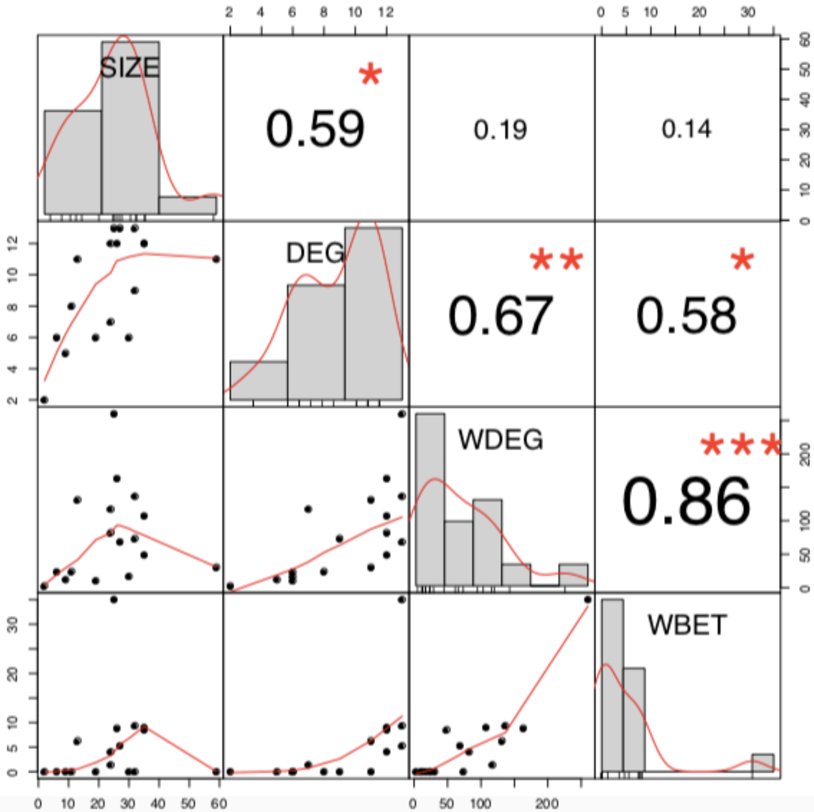
Collaborations among research groups

- The block model formed according to the partition of nodes obtained by the Louvain algorithm
  - #nodes = 17 (research groups)
  - #links = 79 (collaborations between research groups)

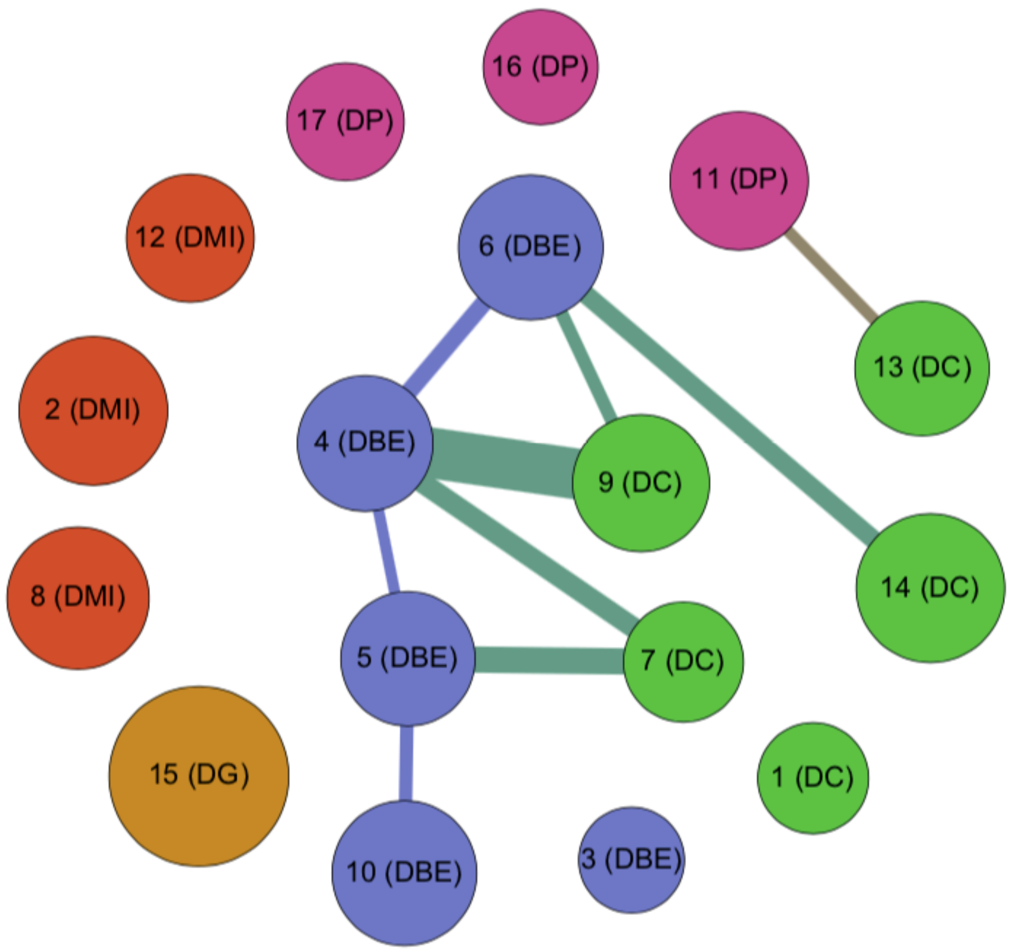


- **Expected:** groups that have strong collaborations tend to be more institutionally important
- **The importance of research groups and the strength of inter-group research collaboration are independent of group size**

Fig. 8.6: The Spearman correlation matrix of the size, degree (DEG), weighted degree (WDEG) and weighted betweenness centrality (WBET) of nodes in the collaboration network of FS-UNS research groups.



- 8 isolated research groups in the reduced collaboration network of FS-UNS research groups containing only strong links
- Research groups from DMI, DG and DP exhibit a weak degree of inter-group research collaborations



- **Researchers involved in inter-group research collaborations are significantly more productive, collaborative and institutionally important**

Metric	Avg(G_1)	Avg(G_2)	U	p	NHA	PS_1	PS_2
PRON	98.6367	26.8662	8221.5	$< 10^{-4}$	no	0.78	0.21
PROF	26.5373	11.1954	11043	$< 10^{-4}$	no	0.71	0.29
PROS	27.8801	11.6127	11325	$< 10^{-4}$	no	0.69	0.28
SRCI	151.3667	51.1648	9037.5	$< 10^{-4}$	no	0.76	0.24
COLL	65.1873	17.5845	5301	$< 10^{-4}$	no	0.86	0.14
LCOLL	17.4569	7.4014	6744.5	$< 10^{-4}$	no	0.81	0.16
ECOLL	47.7303	10.1831	5932	$< 10^{-4}$	no	0.84	0.15
WCOLL	95.1948	23.0563	7651.5	$< 10^{-4}$	no	0.80	0.20
WLCOLL	48.7996	14.9347	8520.5	$< 10^{-4}$	no	0.78	0.22
WECOLL	46.3951	8.1216	8065	$< 10^{-4}$	no	0.79	0.21
IntraDEG	10.3408	7.4014	12475.5	$< 10^{-4}$	no	0.64	0.30
InterDEG	7.1161	0.0000	0	$< 10^{-4}$	no	1.00	0.00
WIntraDEG	43.8952	14.9347	9557	$< 10^{-4}$	no	0.75	0.25
WInterDEG	4.9045	0.0000	0	$< 10^{-4}$	no	1.00	0.00
BET	687.4335	73.2130	5683	$< 10^{-4}$	no	0.84	0.14
CLO	0.3291	0.2676	3142	$< 10^{-4}$	no	0.92	0.08
EVC	0.0031	0.0013	6730	$< 10^{-4}$	no	0.82	0.18
WBET	457.4900	95.9977	12196.5	$< 10^{-4}$	no	0.51	0.16
WCLO	0.6060	0.4651	8647.5	$< 10^{-4}$	no	0.77	0.23
CC	0.4866	0.6553	11824	$< 10^{-4}$	no	0.30	0.68

- Comparison of research groups: analyzed group superiority graphs corresponding to productivity and collaboration metrics
- **PRON and SRCI — biased measures of research productivity**

Metric	Nodes	Links	Superior groups	Inferior groups	Bipartite structure
PRON	7	7	3	4	yes
PROF	0	0	/	/	/
PROS	0	0	/	/	/
SRCI	11	10	2	9	yes
COLL	13	24	9	4	yes
WCOLL	10	12	4	6	yes

Fig. 8.8: The group superiority graph corresponding to the PRON research productivity metric.

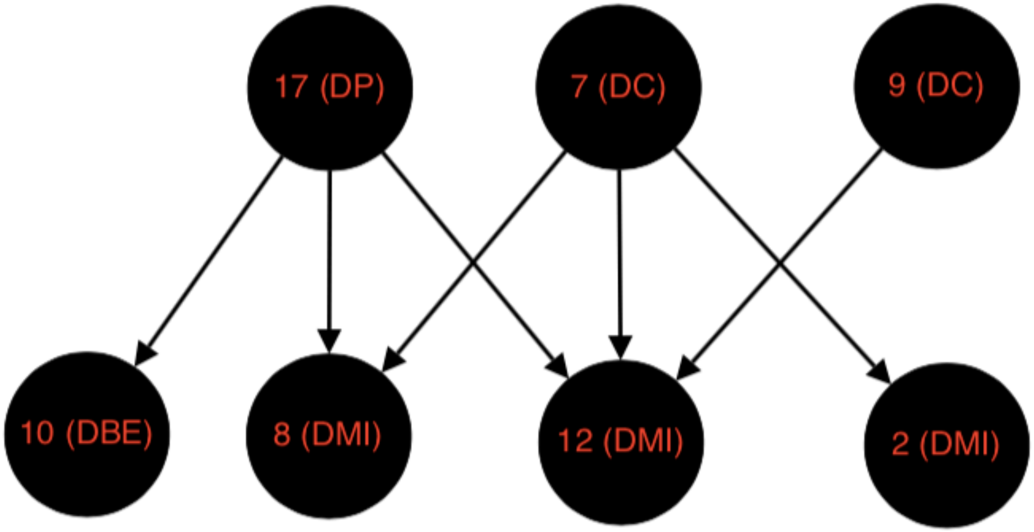


Table 8.14: The links in the group superiority graphs corresponding to the COLL and WCOLL research collaboration metrics.

Metric	Superior group	Superior to
COLL	C17 (DP)	C2 (DMI), C12 (DMI), C8 (DMI), C10 (DBE)
	C11 (DP)	C2 (DMI), C12 (DMI), C8 (DMI)
	C7 (DC)	C2 (DMI), C12 (DMI), C8 (DMI)
	C9 (DC)	C2 (DMI), C12 (DMI), C8 (DMI)
	C14 (DC)	C2 (DMI), C12 (DMI), C8 (DMI)
	C15 (DG)	C2 (DMI), C12 (DMI), C8 (DMI)
	C5 (DBE)	C2 (DMI), C12 (DMI), C8 (DMI)
	C6 (DBE)	C12 (DMI)
	C13 (DC)	C12 (DMI)
WCOLL	C17 (DP)	C6 (DBE), C15 (DG), C8 (DMI), C2 (DMI), C12 (DMI), C10 (DBE)
	C7 (DC)	C8 (DMI), C2 (DMI), C12 (DMI)
	C9 (DC)	C8 (DMI), C12 (DMI)
	C5 (DBE)	C12 (DMI)

Male vs Female FS-UNS researchers

- No SSD in productivity of male and female FS-UNS researchers
- Male FS-UNS researchers tend to have higher betweenness centrality
  - more important for institutional cohesion at the macro-scale
- Female FS-UNS researchers tend to have higher clustering coefficient
  - more important for institutional cohesion at the micro-scale

Metric	$\langle M \rangle$	$\langle F \rangle$	U	MWU-p	PS_m	PS_f	D	KS-p
PRO-N	82.72	65.40	20741	0.63	0.51	0.48	0.07	0.71
PRO-F	25.52	17.85	19975.5	0.27	0.53	0.47	0.10	0.29
PRO-S	25.83	19.33	20754	0.64	0.50	0.47	0.09	0.40
CI	132.05	101.40	20056	0.30	0.53	0.47	0.10	0.23
DEG	51.24	44.98	20440	0.47	0.52	0.47	0.06	0.78
LDEG	13.91	13.26	21194	0.91	0.49	0.48	0.05	0.95
EDEG	37.33	31.72	20301.5	0.40	0.52	0.47	0.07	0.66
BET	579.38	389.05	18564.5	**0.02**	0.55	0.42	0.14	**0.04**
CLO	0.30	0.31	18666	0.23	0.46	0.53	0.12	0.10
CC	0.49	0.55	18545	**0.02**	0.42	0.55	0.13	**0.04**
CGD	0.39	0.41	19004	0.35	0.45	0.51	0.12	0.10

Outline

- Introduction
- Methodology
- Case study
- Main results
- **Conclusions**

Conclusions

- Institutional research information systems offer great opportunities for analysis of intra-institutional research collaboration
  - Researchers described by institutionally relevant attributes, local research evaluation policies supported, no name disambiguation, etc.

- Methodology to analyze intra-institutional research collaboration based on enriched co-authorship networks with an underlying institutional structure

- Main conclusions from our case study
  - FS-UNS research community is highly cohesive, research departments are nearly Radicchi strong clustering, strong community structure
  - Researchers from the network core, researchers involved in inter-department collaborations and researchers involved in inter-group collaborations tend to be more productive, collaborative and institutionally important
  - SRCI should be avoided in strategic and administrative decision making

References

M. Savić, M. Ivanović, B. Dimić Surla. Analysis of intra-institutional research collaboration: a case of a Serbian faculty of sciences. *Scientometrics* 110(1): 195–216, 2017.

M. Savić, M. Ivanović, M. Radovanović, B. Dimić Surla. Towards Culture-Sensitive Extensions of CRISs: Gender-Based Researcher Evaluation. In *Proc. of MEDI'16*, 332-345, Almería, Spain, 2016.

Chapter 8 in "Complex Networks in Software, Knowledge and Social Systems" (to appear)