Structure and dynamics of scientific collaboration networks of Indian researchers in APS publications

THESIS DEFENSE SEMINAR

Chakresh Kumar Singh | 14310048 | 16th April 2021

Final Defense Committee Members:

Thesis Advisors:

Prof. Shivakumar Jolad
Prof. Krishna Kanti Dey

DSC Members:

Prof. Anand Sengupta
Prof. Anirban Dasgupta

IITGN:

Prof. Anand Sengupta
Prof. Anirban Dasgupta
Prof. B. Prasanna Venkatesh
Prof. Neeldhara Misra
Prof. Shivakumar Jolad
Prof. Krishna Kanti Dey

External:

Prof. Abyayananda Maiti (IIT Patna)

Overview :

- Motivation
- Where this work contributes?

- Evolution of the Indian Co-authorship Networks in APS
- Indian Institutes Collaboration Networks
- Correlation between Citation and Co-authorship Networks

- Where this leads?

Overview :

- Motivation
- Where this work contributes?

This Thesis

- Evolution of the Indian Co-authorship Networks in APS
- Indian Institutes Collaboration Networks
- Correlation between Citation and Co-authorship Networks

- Where this leads?

Motivation

- Kuhn's Idea of Structure of Scientific Revolutions
    - Changes in Perceptions - "Paradigm Shifts"
    - Interaction between the community of researchers - "Collaboration"
- Quantitative assessment has been abundant. However, less work on linking the differences in interactions across different domains, regions etc.
- Scattered literature focussing on small communities especially in the Indian context

Where this work contributes?

- In-depth analysis tracing and correlating the collaboration patterns for a closed community (otherwise hidden in macroscopic analyses)

- A small step forward towards linking the social capital of researchers i.e their collaborations with the evolution of scientific domains

Major research domains in each APS Journal

- PRA (Physical Review A): Atomic and Molecular Physics
- PRB (Physical Review B): Condensed Matter and Material Physics
- PRC (Physical Review C): Nuclear Physics
- PRD (Physical Review D): Particle, Fields, Gravitation, and Cosmology
- PRE (Physical Review E): Statistical, Non-linear, Biological, and Soft Matter Physics
- PRL (Physical Review L): Applied, Fundamental and Interdisciplinary Physics.

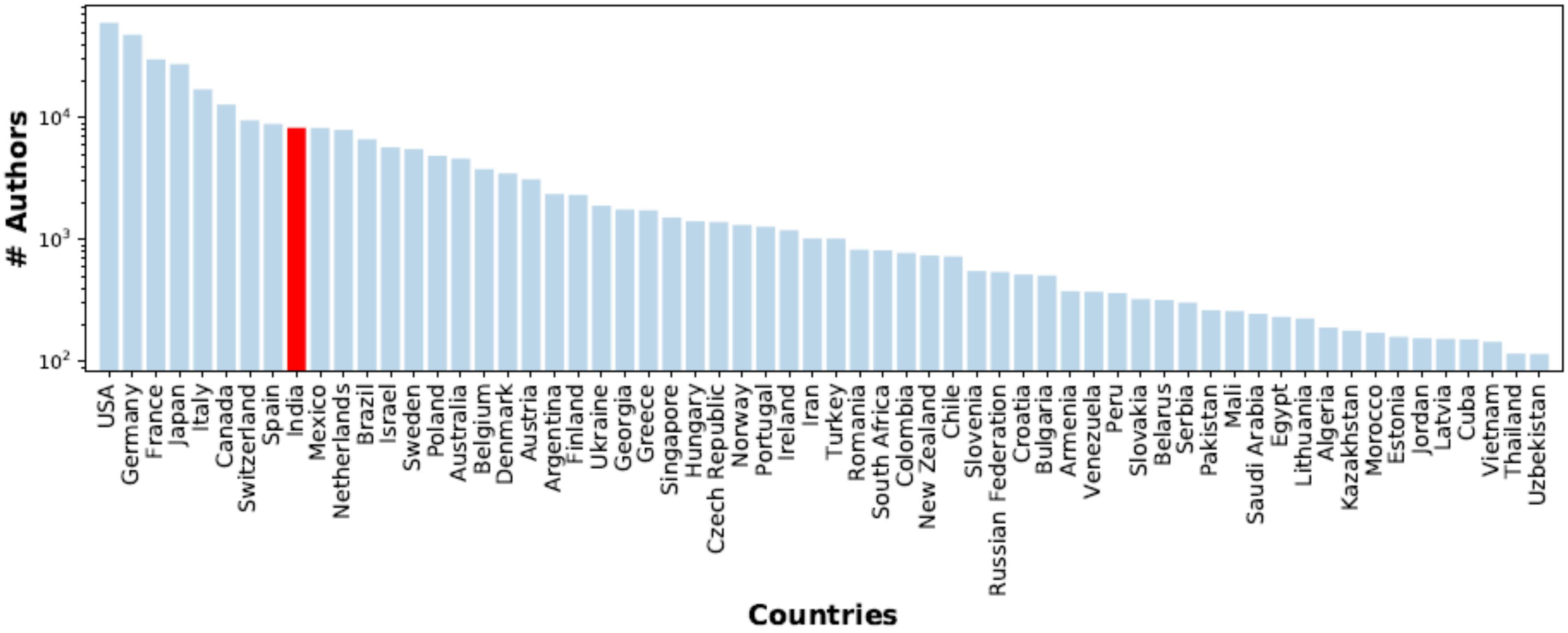
Data-Set

- American Physical Society Publications (APS) from 1905-2013
- Indian papers identified using the affiliation information
- Analysis was focused between time periods 1970-2013
- Identified unique Indian authors (8084)
- Identified unique institutions (677)

Publications counted between 1970-2013

	Papers		
Journals	Total	Indian	Percent
PRA	65,170	2,146	3.29%
PRB	161,257	3,941	2.44%
PRC	34,443	1,967	5.71%
PRD	69,481	3,197	4.60%
PRE	46,009	1,658	3.60%
PRL	110,080	1,794	1.63%
TOTAL	486,440	14,703	3.02%

Total Authors in APS Publications



Naming Disambiguation - Indian authors

Step 1 - **Set** of all Names

Step 2 - Correcting for abbreviated forms

Step 3 - correcting for spelling error

A. Gupta, Ajay Gupta, S. Dinesh, A. Bagchi, A. Saha, B. Das, ….. C. Sigh

Combine names that satisfy:

(same letter for the first name) & (same last name)

In the reduced set look for similar strings and if they have a significant overlap in collaborators combine them

We are left with 8084 unique Indian authors

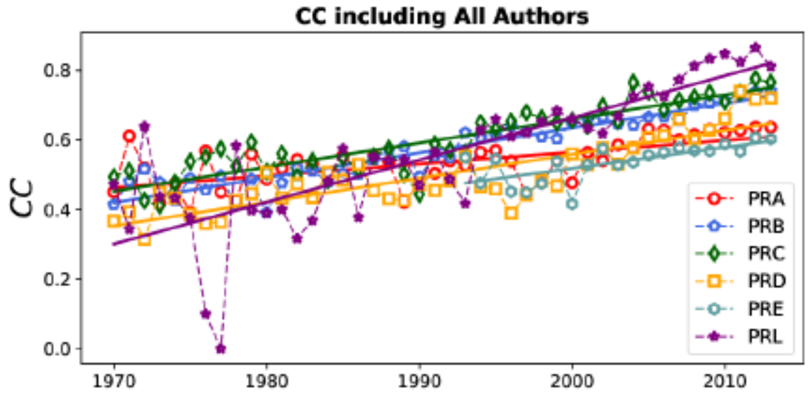
Collaboration Coefficient (CC)

$$CC = 1 - \frac{\sum_{j=1}^{q} \frac{1}{j} f_j}{N}$$

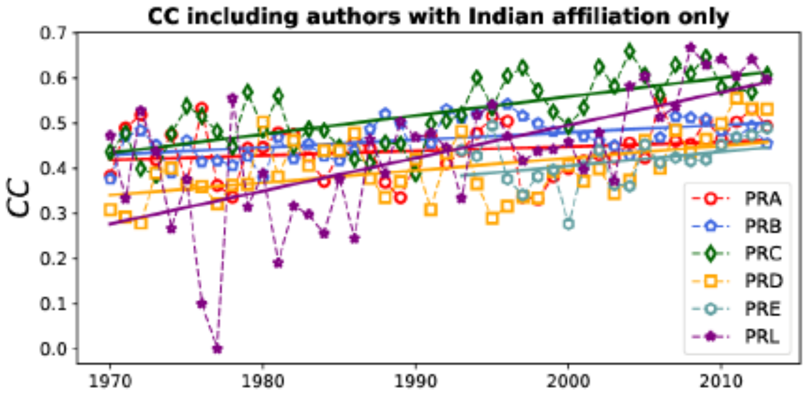
N : Number of Papers in an year

f_j : Number of j authored papers

q : maximum authors in a paper in a given year



(a) Indian-Indian and Indian-Foreign

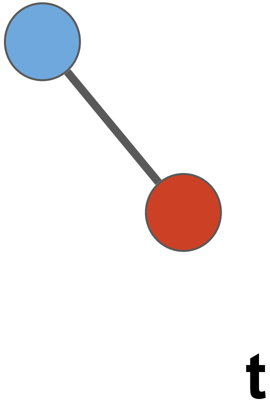


(b) Indian-Indian

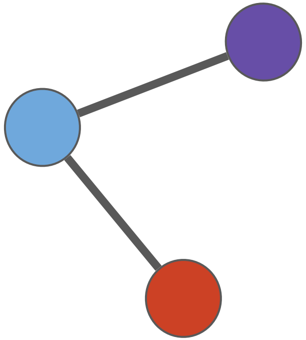
	Collaboration coefficient gradient ($10^{-3}/yr$)		
Journal	**I-I**	**I-F**	**I-IF**
PRA	0.99	2.91	3.39
PRB	1.65	8.16	7.13
PRC	4.17	7.42	6.92
PRD	2.73	8.4	6.83
PRE	3	4	6
PRL	7.29	15.69	12.09

- **PRL has the highest CC**
- **PRA has the lowest CC**
- **CC is more with respect to foreign authors than with Indians**

Co-authorship Networks

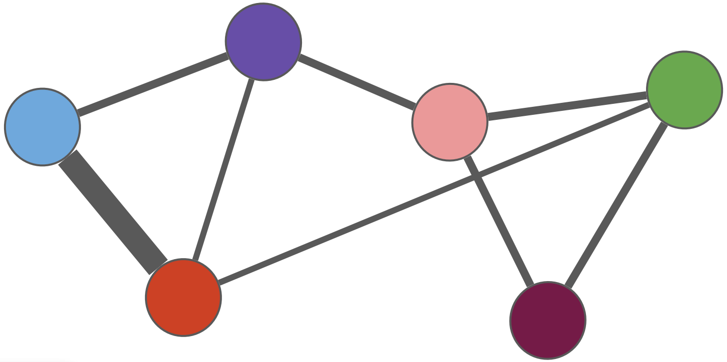


Co-authorship Networks



t+1

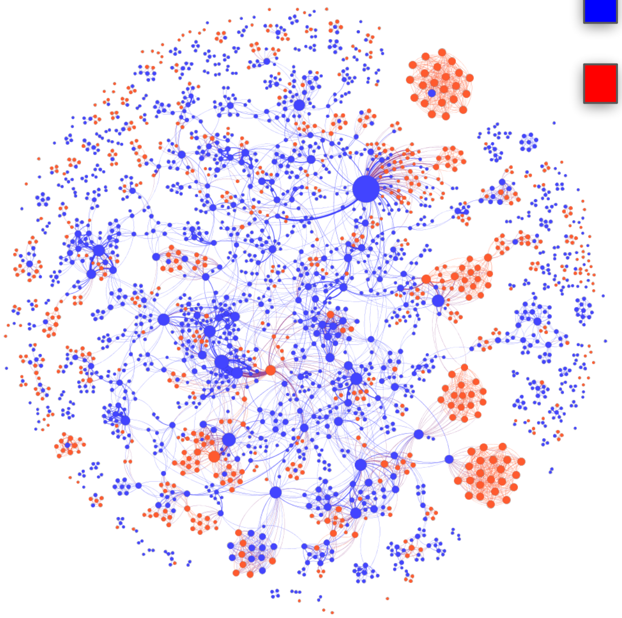
Co-authorship Networks



Different growth mechanisms can lead to different networks

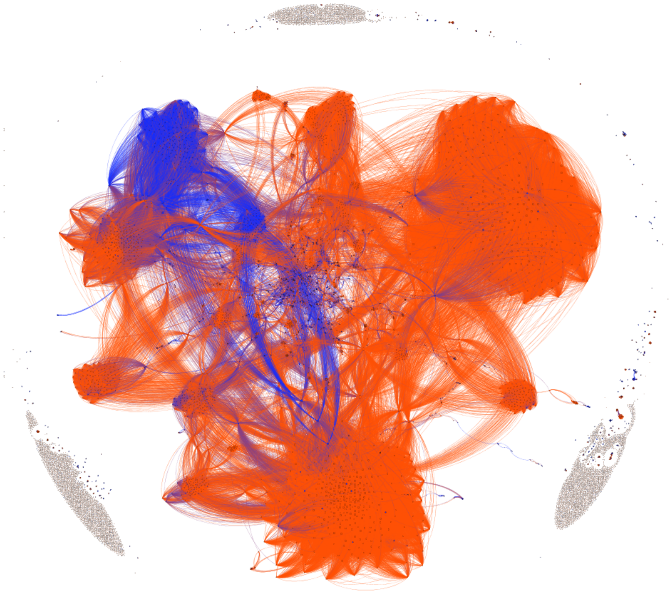
t+2

Co-authorship Networks



Indian

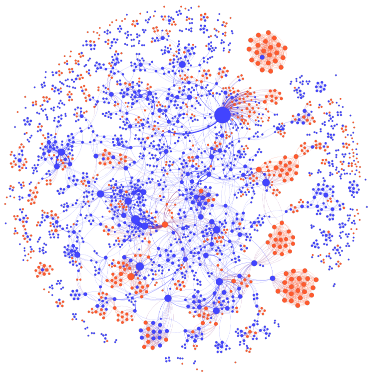
Foreign



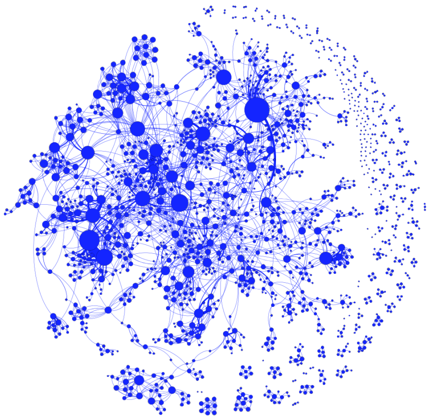
PRA

PRL

Macroscopic Network Properties



Removing Foreign Nodes



	PRA		PRB		PRE		PRD		PRC		PRL	
	I-IF	I-I	I-IF	I-I	I-IF	I-I	I-IF	I-I	I-IF	I-I	I-IF	I-I
$\langle k \rangle$	5.3	4.1	8.3	6.2	3.9	3.1	33.07	8.4	31	7.14	27.17	13.96
$\langle C \rangle$	0.8	0.76	0.82	0.74	0.82	0.77	0.82	0.72	0.85	0.76	0.9	0.84
$\langle l \rangle$	6.4	6.31	4.5	4.5	6.75	6.64	4.37	6.49	3.26	3.65	3.98	5.08

Table 2.4. Network statistics- average degree $\langle k \rangle$, clustering coefficient $\langle C \rangle$ and mean path length $\langle l \rangle$ of Indian-Indian and Indian-foreign coauthorship network for APS journals from 1970-2013.

Structure of Co-authorship Networks

Power Law in Degree Distribution

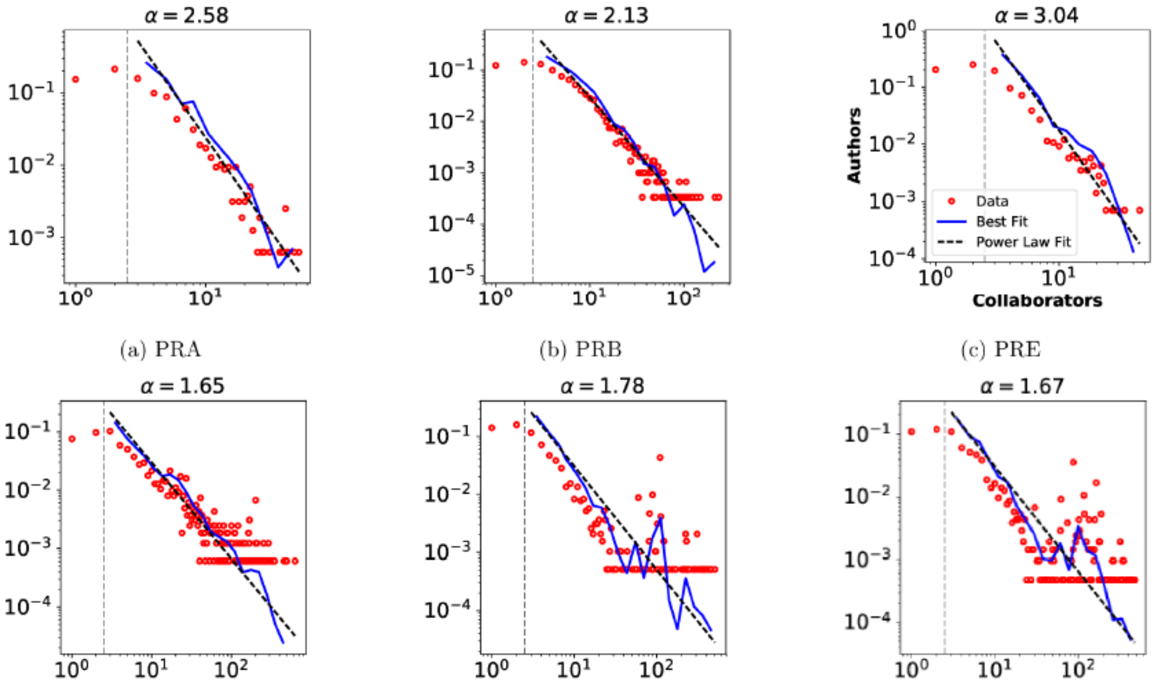
$$p(k) = Ck^{-\alpha}$$

- The network structure and dynamics processes on it significantly differ for values of alpha <= 2 vs alpha > 2
    - When alpha <= 2, Link grow faster than nodes, high clustering,
    - When alpha > 2, finite avg. degree.

$$p(k) = \frac{\alpha-1}{k_{min}}\left(\frac{k}{k_{min}}\right)^{-\alpha}$$

$$\langle k \rangle = \int_{x_{min}}^{\infty} kp(k)dk$$

$$\langle k \rangle = \frac{C}{2-\alpha}\left[k^{-\alpha+2}\right]_{k_{min}}^{\infty}$$



(a) PRA (b) PRB (c) PRE

(d) PRC (e) PRD (f) PRL

PRA, PRB and, PRE differ in their alpha values from PRC, PRD and, PRL.

This indicates a strong structural difference between these networks

Centrality measures

Betweenness

$$C_B = \sum_{jk} \frac{n^i_{jk}}{g_{jk}}$$

g_{jk} : No. of shortest paths between j and k

n^i_{jk} : No. of shortest paths between j and k passing through i

Closeness

$$C_i = \frac{n-1}{\sum_{j=1}^{n-1} d(i,j)}$$

n : No. of nodes in the network

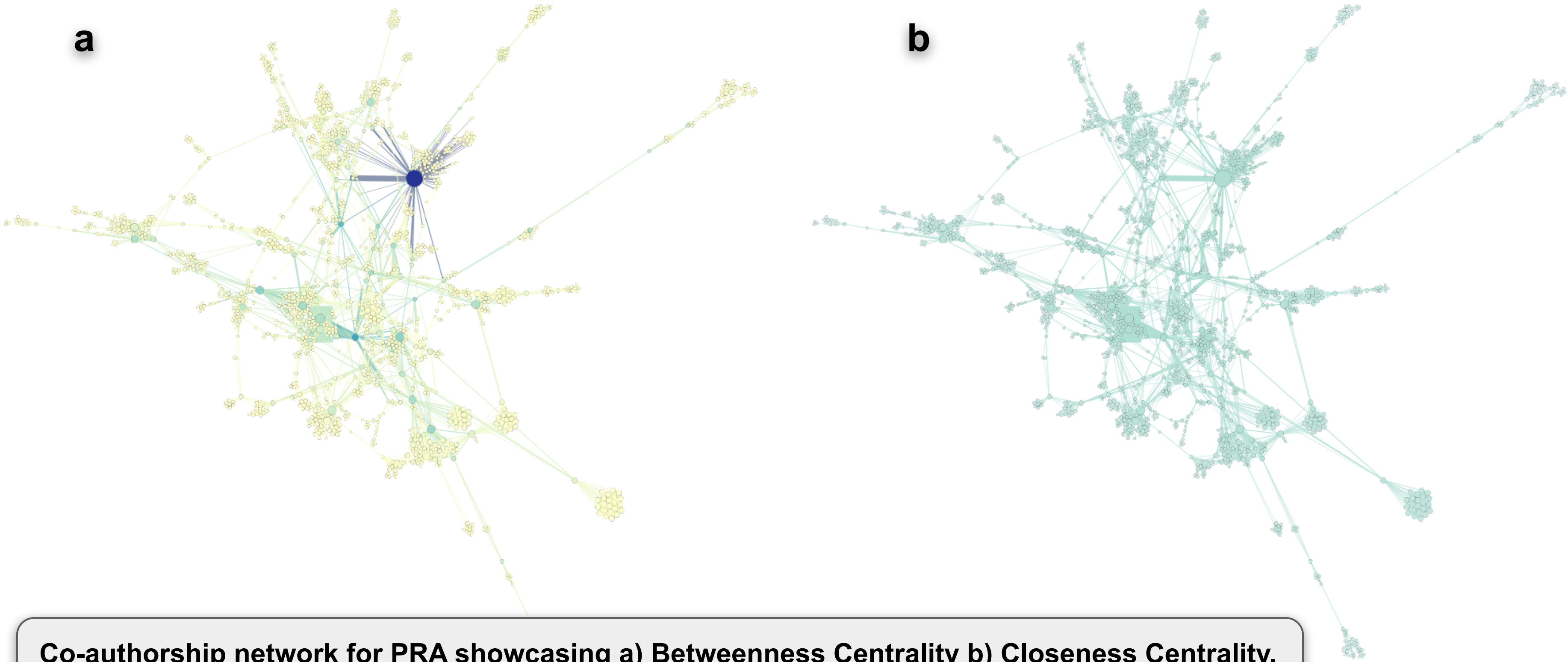
$d(i,j)$: length of shortest path btw. i and j

Coreness

$$C_{nc}(v) = \sum_{w \in N(v)} ks(w)$$

$N(v)$: Set of neighbors of node v

$ks(w)$: k-shell index of neighbor w



Co-authorship network for PRA showcasing a) Betweenness Centrality b) Closeness Centrality. Node size represents degree. Node color represents centrality value. Low (light) → High (dark)

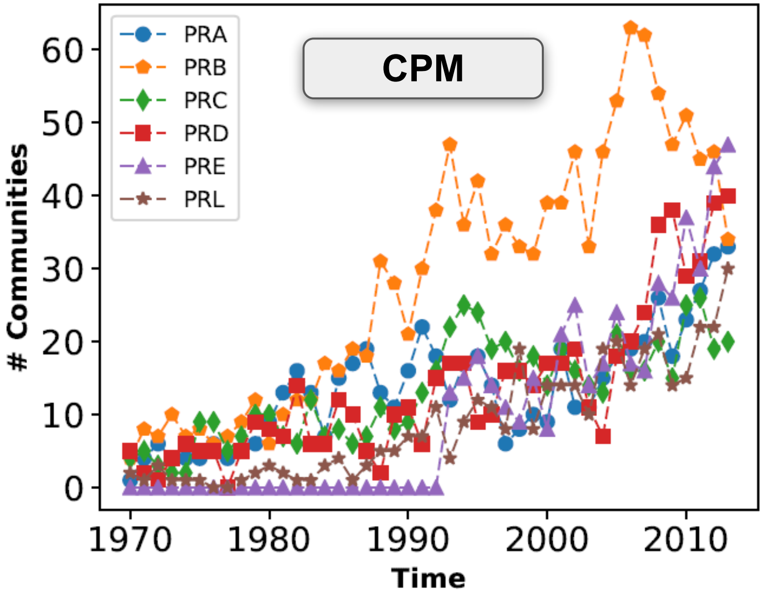
Influential Authors

Step 1	Step 2	Step 3
Rank all authors based on their network centrality values	Select Top 20% authors in the different centrality measures calculated on the network	Authors belonging to the intersection of the sets from Step 2 are considered influential

Influential Indian authors by various centrality measures								
1970- 75	**1976-80**	**1981-85**	**1986-90**	**1991-95**	**1996-2000**	**2001-05**	**2006-09**	**2010-13**
				PRA				
S. C. Mukherjee K. N. Pathak R. Prasad K. Roy S. C. Deorani	S. C. Mukherjee B. C. Saha M. K. Srivastava S. K. Sur G. S. Agarwal	B. Sanjeevaiah A. N. Tripathi S. C. Mukherjee Shyamal Datta S. K. Sur	V. Lakshminarayana V. Gopalakrishna R. Srivastava S. Bhuloka Reddy V. Radha Krishna Murty	M. M. Panja R. R. Puri M. Azam G. S. Agarwal C. L. Mehta	M. B. Kurup K. G. Prasad M. Sarkar K. Vijayalakshmi H. C. Padhi	D. G. Kanhere Anil Kumar P. C. Deshmukh	H. Singhal R. K. Chaudhuri D. Mukherjee R. A. Khan Rajat K. Chaudhuri	H. Singhal J. A. Chakera L. C. Tribedi T. S. Mahesh Ujjwal Sen
				PRB				
Chanchal K. Majumdar G. Rama Rao M. P. Verma S. K. Joshi C. N. R. Rao	T. K. Saxena M. P. Verma K. N. Pathak O. P. Sharma S. K. Joshi	A. N. Basu Vijay A. Singh E. V. Sampathkumaran S. Prakash B. D. Padalia	S. K. Dhar B. D. Paladin S. M. Kanetkar L. C. Gupta J. V. Yakhmi	S. K. Dhar R. Vijayaraghavan Z. Hossain H. R. Krishnamurthy A. S. Tamhane	U. V. Varadaraju B. K. Godwal R. Vijayaraghavan Z. Hossain H. R. Krishnamurthy	S. K. Dhar A. K. Tyagi S. Rayaprol Amish G. Joshi A. K. Sood	B. L. Ahuja S. K. Dhar S. S. Banerjee Sugata Ray Ranjan Mittal	Chandrabhas Narayana S. Singh S. K. Dhar Sugata Ray A. K. Tyagi
				PRC				
M. K. Mehta S. L. Gupta S. Kailas S. C. Pancholi	S. B. Manohar Ashok Kumar S. Kailas Satya Prakash Y. P. Viyogi	V. K. Mittal P. Singh Gulzar Singh Satya Prakash M. V. Ramaniah	M. Saha Y. P. Viyogi A. Goswami	A. Chatterjee Y. P. Viyogi C. Bhattacharya P. Singh S. Kailas	A. Chatterjee S. K. Gupta R. Dutt K. Kar C. Das	D. S. Mukhopadhyay S. Chattopadhyay M. R. Dutta Majumdar Y. P. Viyogi N. K. Rao	B. J. Roy R. Palit V. Jha S. Mukhopadhyay A. Saxena	B. J. Roy R. Palit V. Jha S. Santra S. Mukhopadhyay
				PRD				
N. Panchapakesan G. Rajasekaran K. Datta A. N. Mitra L. K. Pandit	N. S. Arya T. R. Govindarajan J. N. Misra A. Mozumder S. Roy	G. Singh N. Mukunda Ashok Goyal S. N. Biswas V. K. Gupta	I. S. Mittra M. Kaur N. K. Rao V. K. Gupta	J. B. Singh J. M. Kohli S. K. Gupta N. K. Rao I. S. Mittra	D. P. Roy H. Mendez V. Kapoor B. Choudhary	Probir Roy M. Sami Ashok Goyal Soumitra SenGupta Biswarup Mukhopadhyaya	Probir Roy Nilmani Mathur Ritesh K. Singh J. B. Singh Manmohan Gupta	B. Bhuyan Anirban Kundu Monoranjan Guchait R. Sinha Debajyoti Choudhury
				PRE				
				S. R. Sharma M. Saha Ramakrishna Ramaswamy A. Sen P. Nandy	Arnab Majumdar Debashish Chowdhury P. S. Goyal S. Lakshmibala Sriram Ramaswamy	Moumita Das S. Mitra S. S. Manna C. Dasgupta Ramakrishna Ramaswamy	S. Kumar Sudeshna Sinha M. Lakshmanan A. K. Sood D. V. Senthilkumar	Subir K. Das Sanjay Puri Dipankar Bandyopadhyay M. Lakshmanan B. Kundu
				PRL				
Virendra Singh K. Govindarajan P. Chandra Sekhar	R. Ramachandran A. N. Mitra	I. S. Mittra R. Joseph S. K. Tuli Y. Prakash S. Satti	I. S. Mittra S. K. Badyal M. M. Aggarwal N. K. Rao V. K. Gupta	R. Raniwala S. Kachroo	D. S. Mukhopadhyay S. K. Badyal M. D. Trivedi S. S. Sambyal M. R. Dutta Majumdar	P. K. Kaw A. Kumar S. S. Sambyal P. Raychaudhuri Somendra M. Bhattacharjee	Bijaya K. Sahoo S. P. Singh T. Senthil S. S. Sambyal P. K. Kaw	S. P. Behera A. D. Ayangeakaa L. C. Tribedi G. Ravindra Kumar S. S. Ghugre

Community Dynamics

- What is the structure of co-authorship communities in Indian researchers ?
- How do they evolve with time ?



- Implications ??

[Clique Percolation Method (CPM)]

Evolution of Communities

- Birth
- Growth
- Splitting
- Merging
- Death

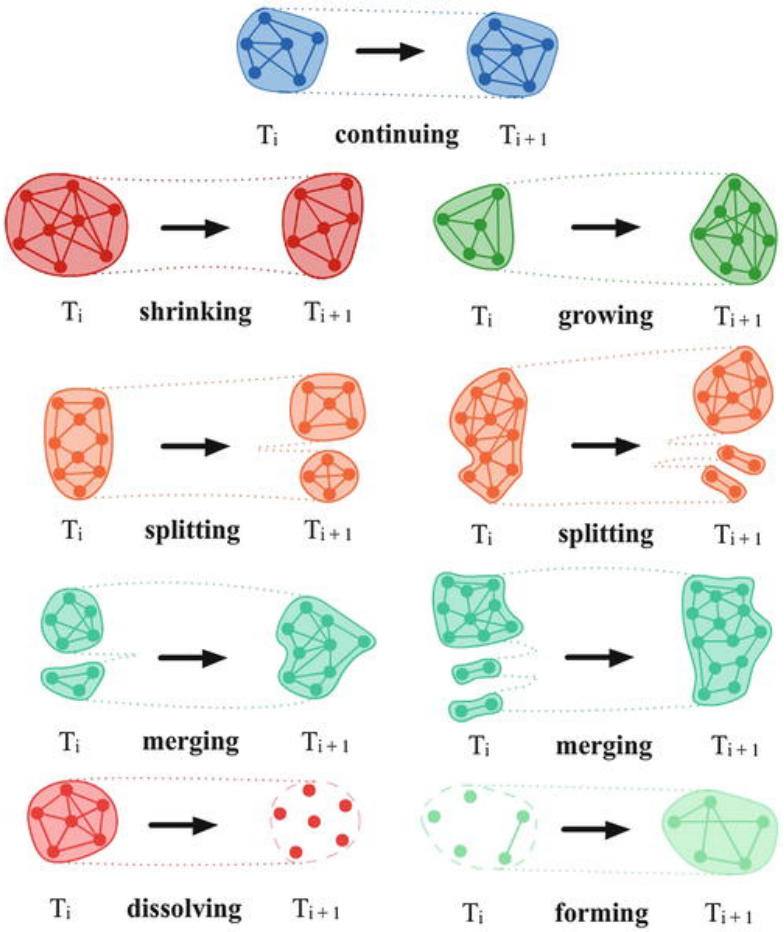


Image from: https://arxiv.org/ftp/arxiv/papers/1605/1605.00069.pdf

Tracing the evolution of communities

- **Match the communities at two consecutive time steps i.e for a community ‘A’ at t0 find its best match at t0+t**
- **A community is dead if no match found.**
- **Calculate the overlap at every time step**

Method borrowed from - *[Palla, G., Barabási, AL. & Vicsek, T. Quantifying social group evolution. Nature 446, 664–667 (2007).]*

[Slides for the method]

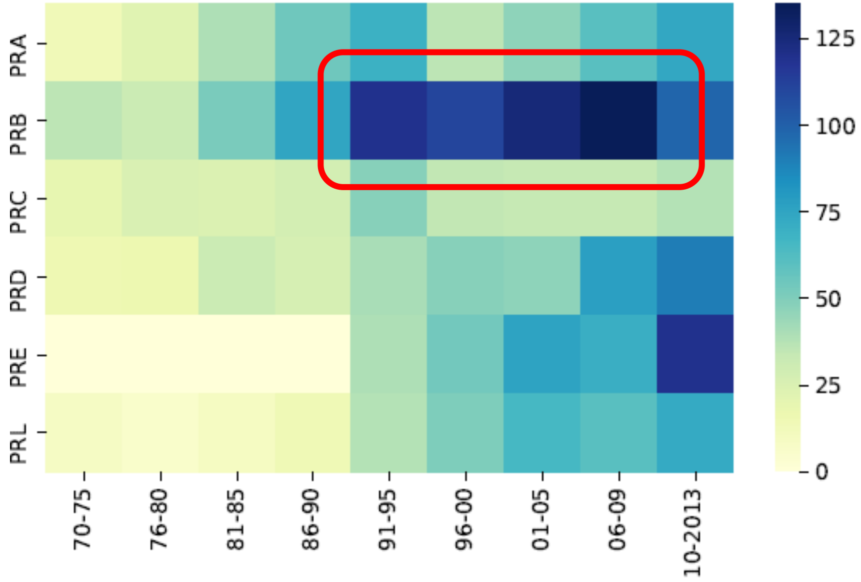
Overlap

$$C^A(t_0, t_0+t) = \frac{|A(t_0) \cap A(t_0+t)|}{|A(t_0) \cup A(t_0+t)|}$$

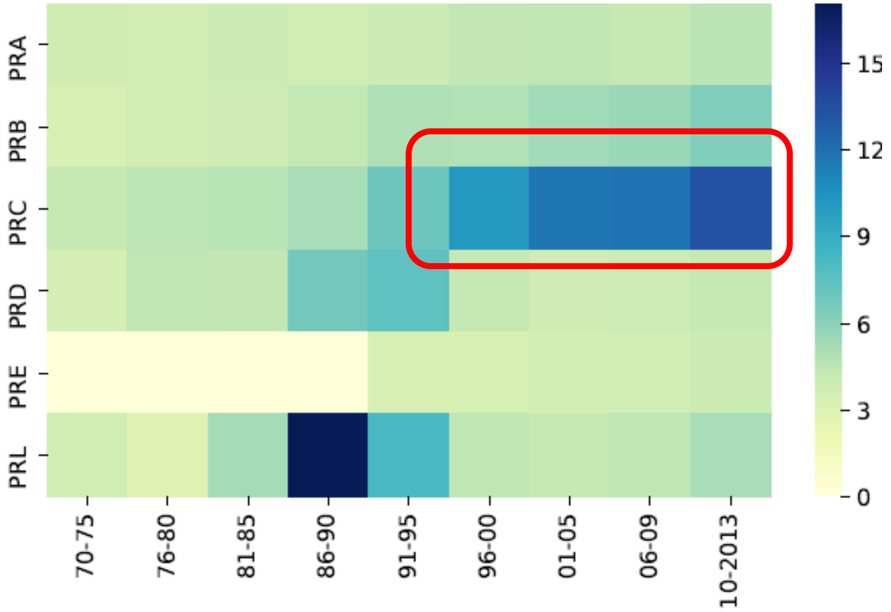
Stationarity

$$\zeta(A) \equiv \frac{\sum_{t=t_0}^{t_{max}-1} C^A(t, t+1)}{t_{max} - t_0 - 1}$$

Aggregated the networks into 9 time periods

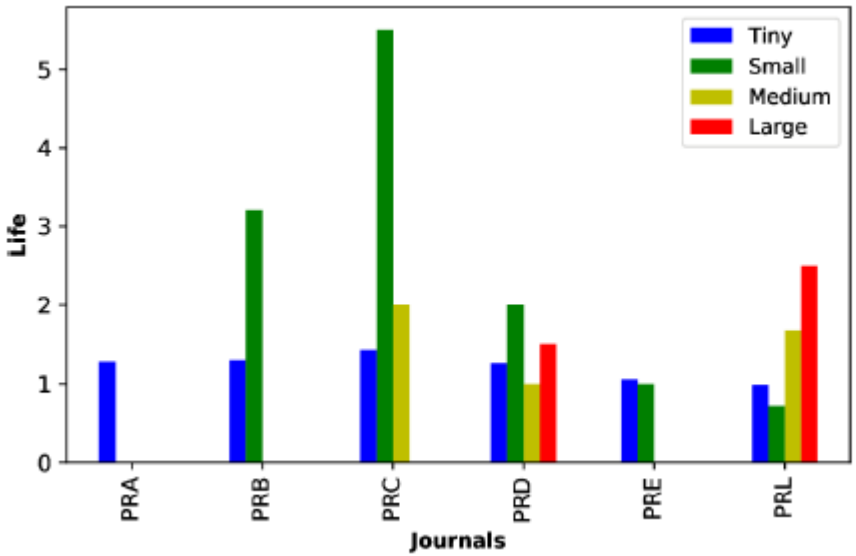


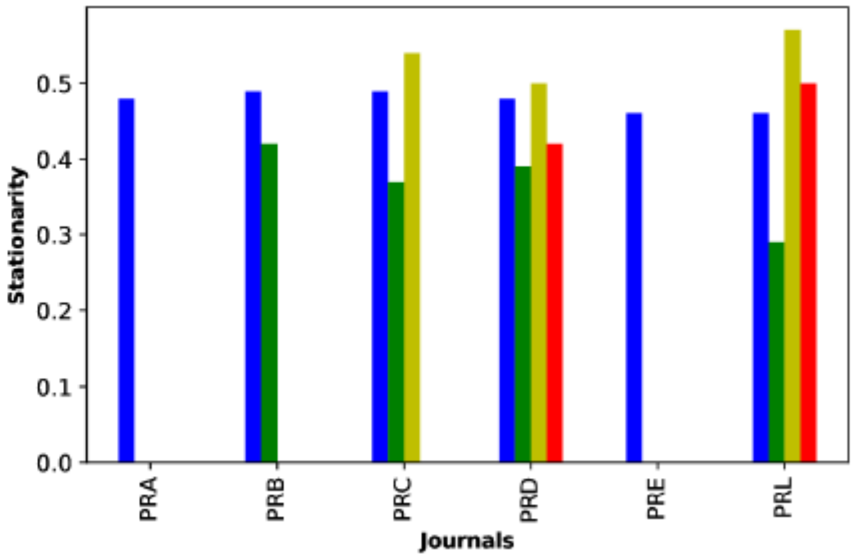
(a) Number



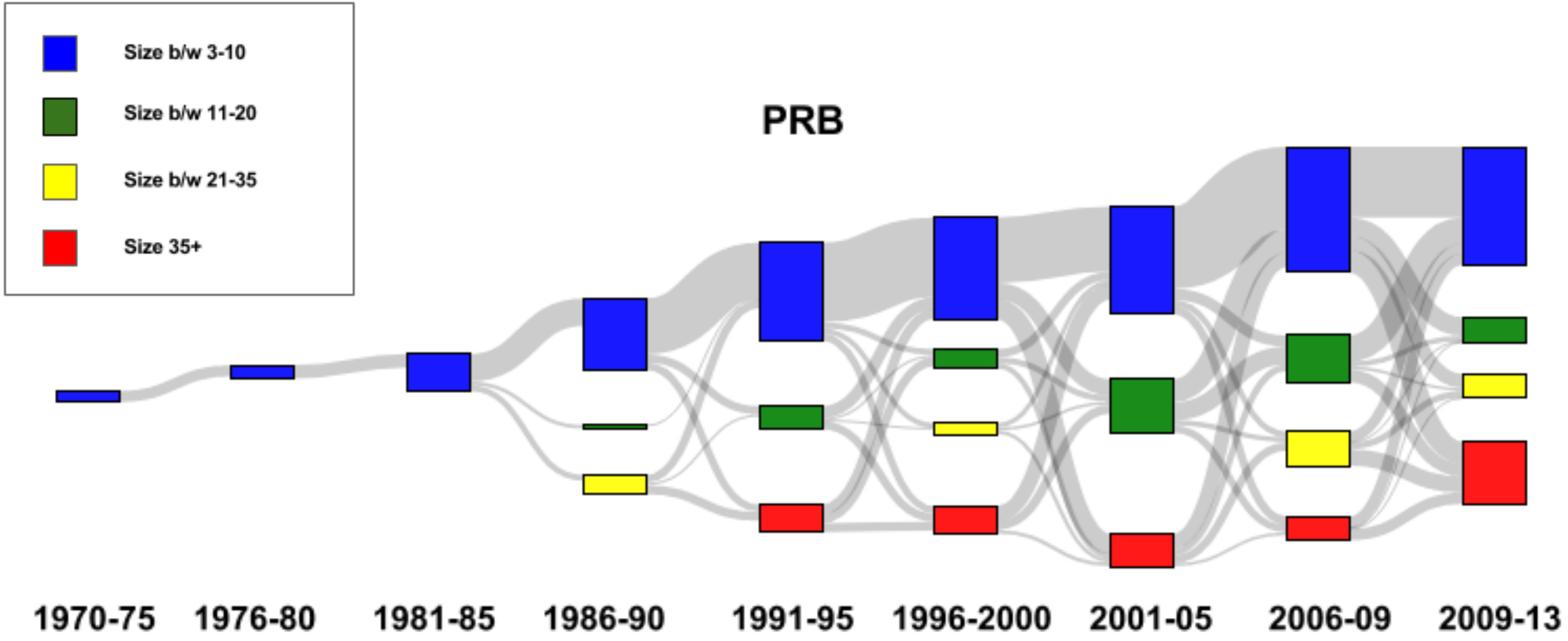
(b) Size

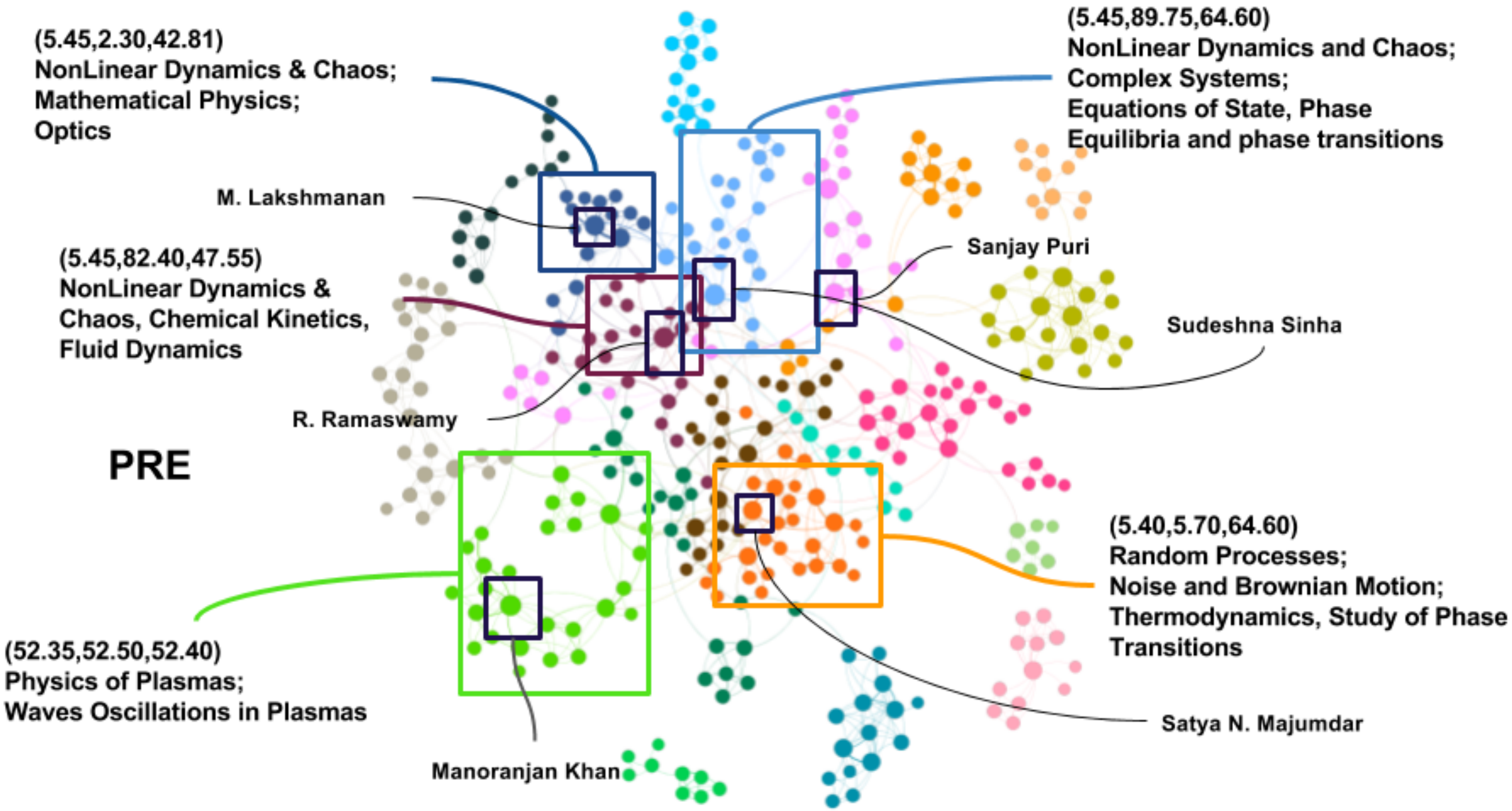
Life time and Stationarity





Visualizing the Dynamics



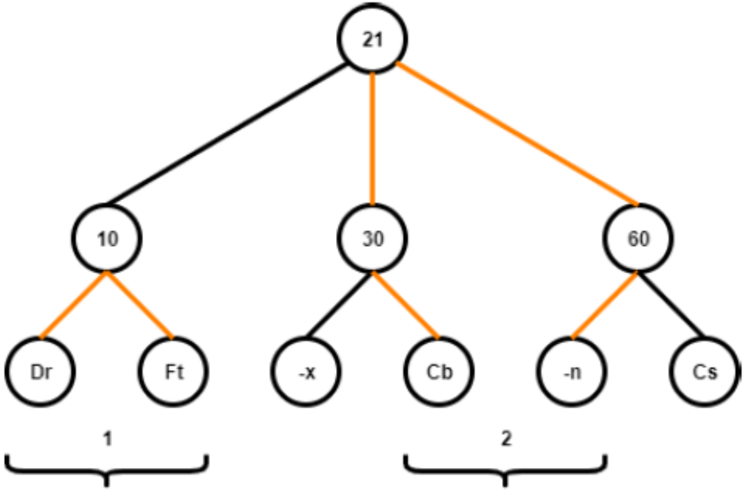

(5.45,2.30,42.81)
NonLinear Dynamics & Chaos;
Mathematical Physics;
Optics
M. Lakshmanan
(5.45,82.40,47.55)
NonLinear Dynamics &
Chaos, Chemical Kinetics,
Fluid Dynamics
R. Ramaswamy
PRE
(52.35,52.50,52.40)
Physics of Plasmas;
Waves Oscillations in Plasmas
Manoranjan Khan
(5.45,89.75,64.60)
NonLinear Dynamics and Chaos;
Complex Systems;
Equations of State, Phase
Equilibria and phase transitions
Sanjay Puri
Sudeshna Sinha
(5.40,5.70,64.60)
Random Processes;
Noise and Brownian Motion;
Thermodynamics, Study of Phase
Transitions
Satya N. Majumdar

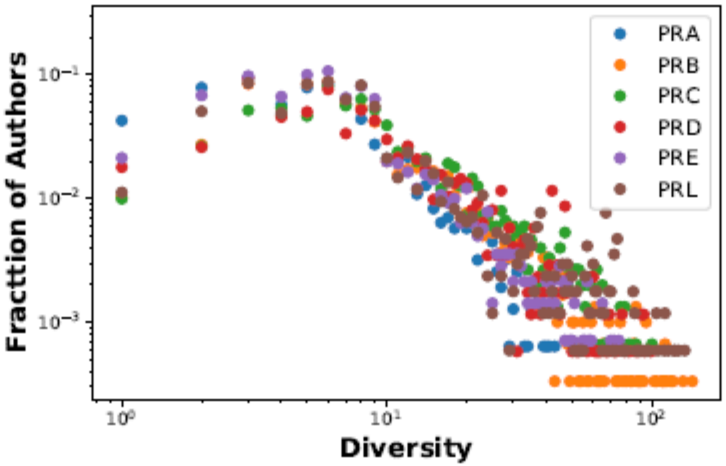
Diversity of authors

- Distance between two PACS is the minimum steps needed to reach a common root

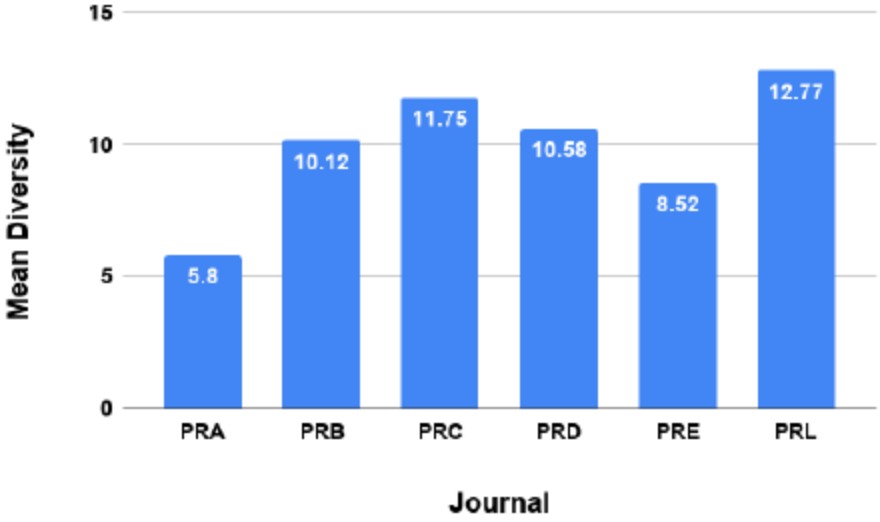
- Then we use Weitzman Diversity measure to define the diversity of a paper

Method borrowed from : Enduri, Murali Krishna, I. Vinod Reddy, and Shivakumar Jolad. "Does diversity of papers affect their citations? Evidence from American Physical Society Journals." *2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE, 2015.





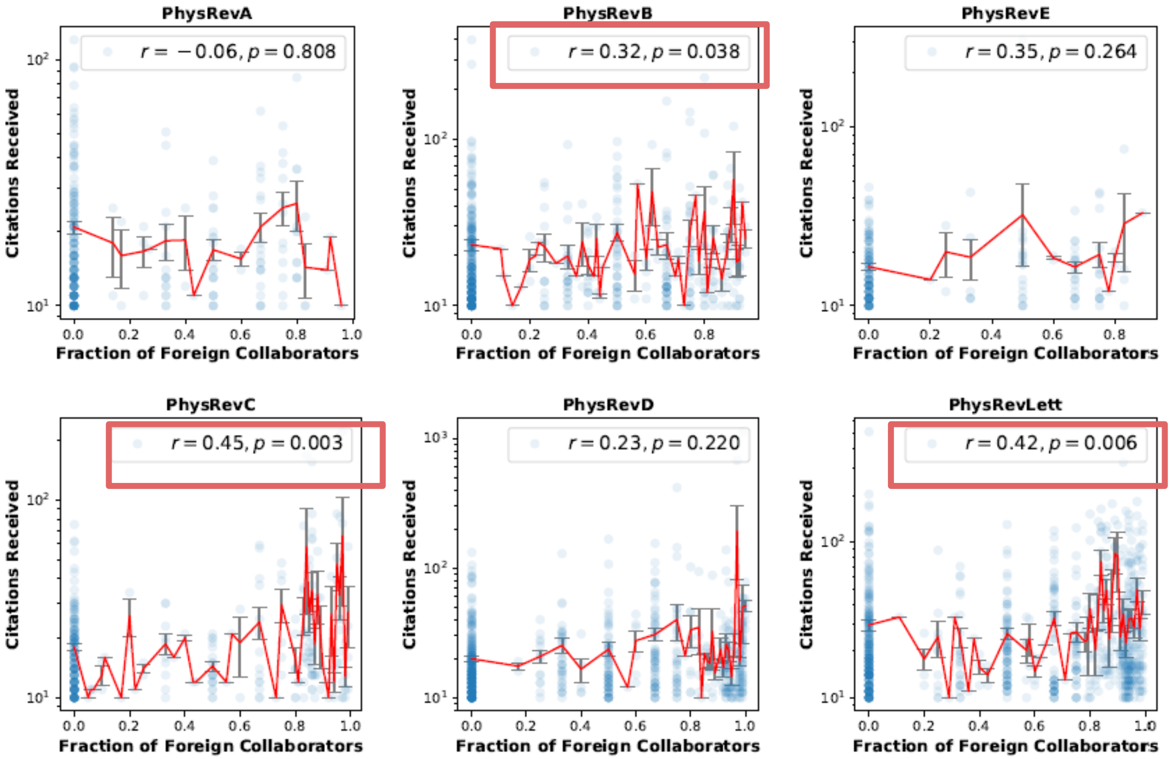
(a)



(b)

(a) Distribution of diversity is similar for all journals. (b) Mean diversity is highest for PRL followed by PRC

Impact of Foreign Collaborations



$$C_F = \frac{(N - I)}{N}$$

N : Total Authors

I : Indian Authors

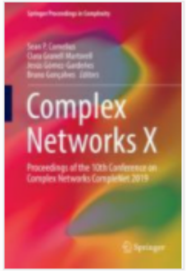
Lower p-values indicate higher reliability in correlation values

Summary of Findings

- Collaboration Structure differs in PRA,PRB and PRE from PRC, PRD and PRL - mainly due to the presence of large experimental collaborations (CERN, LIGO etc.)
- Research in smaller groups in fields of Atomic Physics, Condensed Matter, Non-Linear, Statistical Physics. However, frequent presence of large group collaborations in High Energy Physics, Nuclear Physics
- Foreign Collaboration dominant in PRC, PRD and PRL
- Communities are short living and stationary i.e similar group of authors collaborate for few papers


introduction
abstract
related works
results
novelty
conclusion
Research paper
ilovePhD

Springer Link



International Workshop on Complex Networks
CompleNet 2019: Complex Networks X pp 169-181 | Cite as

Exploring the Role and Nature of Interactions Between Institutes in a Local Affiliation Network

Authors | Authors and affiliations

Chakresh Kumar Singh, Ravi Vishwakarma, Shivakumar Jolad

Conference paper
First Online: 06 March 2019

Part of the Springer Proceedings in Complexity book series (SPCOM)

DOI : https://doi.org/10.1007/978-3-030-14459-3_14

Naming Disambiguation - Indian Institutes

Step 1 - **Set** of all Names

Step 2 - Correcting for different styles of naming

Step 3 - Assigning them into a category

IIT Gandhinagar, I.I.T - Gandhinagar, Indian Institute of Technology Gandhinagar…

7180 names

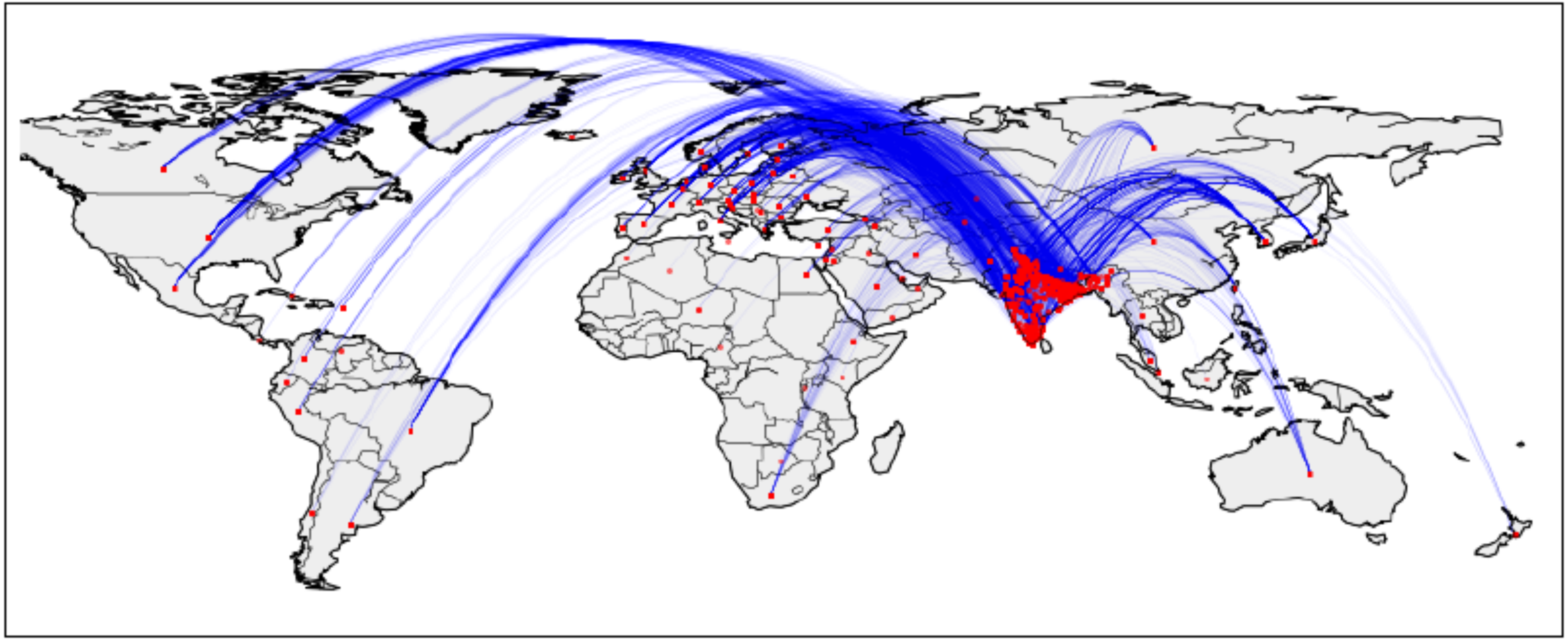
Combine names that satisfy:

(same letter for each word in the title) & (same location - using pincodes)

Manually check the final list and assign each a category based on the UGC classification - special mention : Ravi Vishwakarma

We are left with 677 unique Indian Institutes

Global Collaboration Network

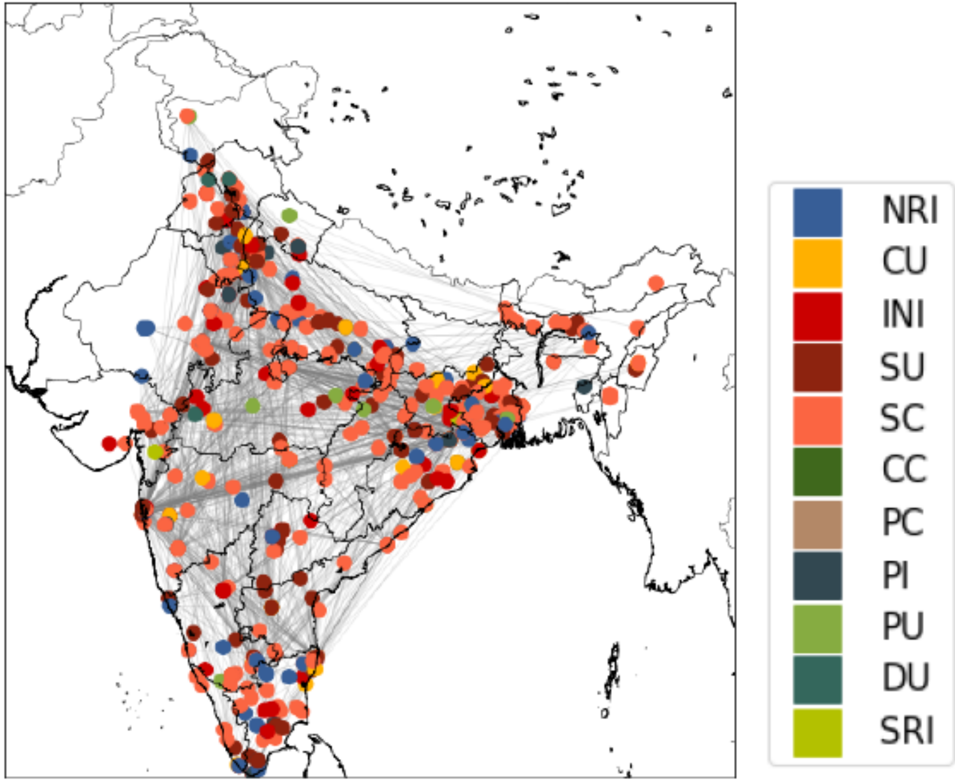


Type of Institutes	Acronym	Function
National Research Institutes	NRI	Research Institutions funded by the central government
Institutes of National Importance	INI	Teaching (both UG and PG) and research Institutions, declared by as INI by Government of India
Central Universities	CU	Public Universities formed by Central Act.
State Universities	SU	Public Universities formed by State Act.
State Colleges	SC	Colleges affiliated to State Universities
Central Colleges	CC	Colleges affiliated to Central Universities
Deemed Universities	DU	Public or Private Universities which can award degrees on their own , and declared as deemed by UGC
Private Universities	PU	Universities established through a state or central act by a sponsoring body
Private Institutes	PI	Stand alone private Institutions recognized by government
State Research Institutes	SRI	Research Institutions funded by the state government

Indian Institutes Collaboration Network

- 993 Uni
- ~10k Standalone Institutes
- ~39k Colleges

Source: [Link]



Each Dot is a Pincode Location

Does Geographic Proximity drive collaboration

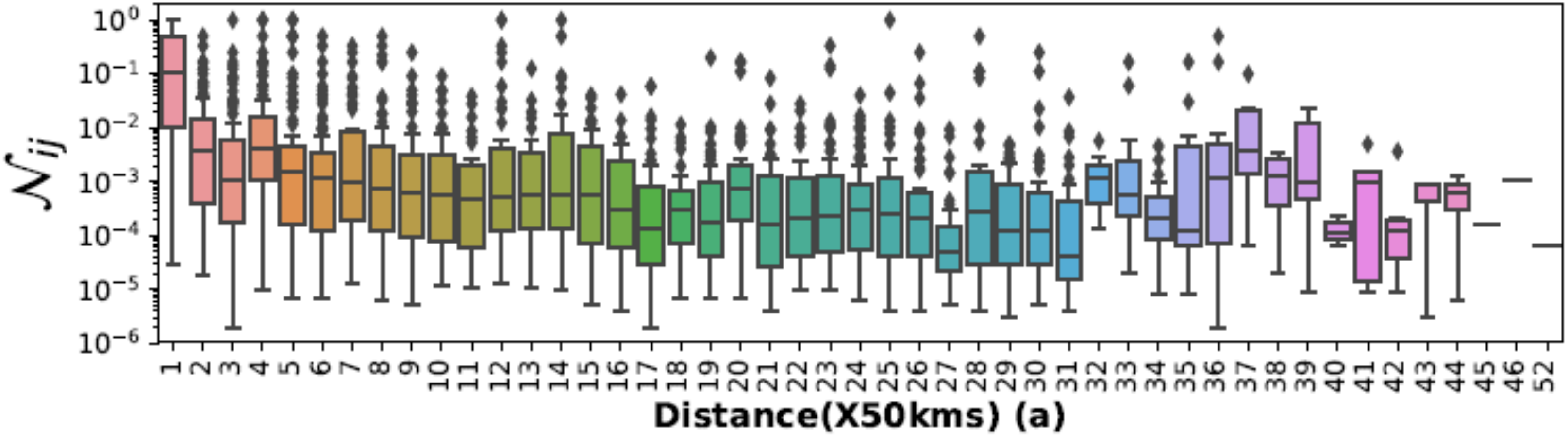
Strength of Collaboration

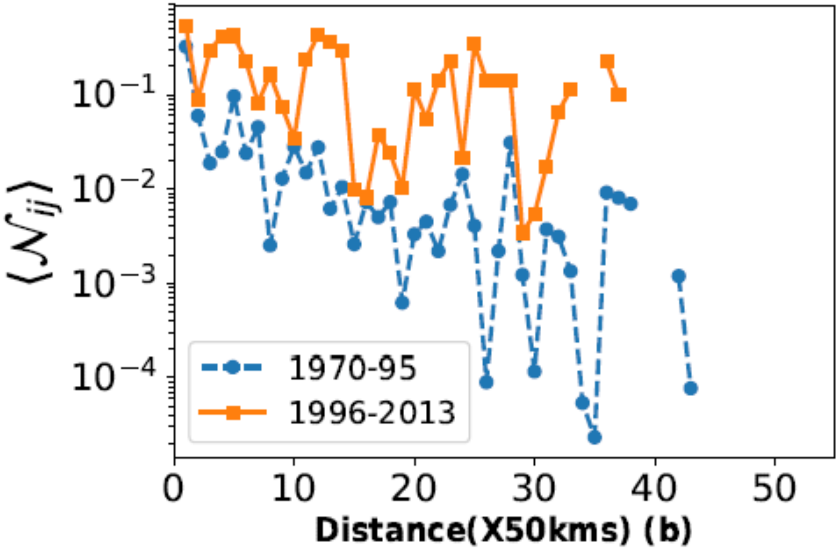
$$\mathcal{N}_{ij} = \frac{\mathcal{P}_{ij}}{w_i \times w_j}$$

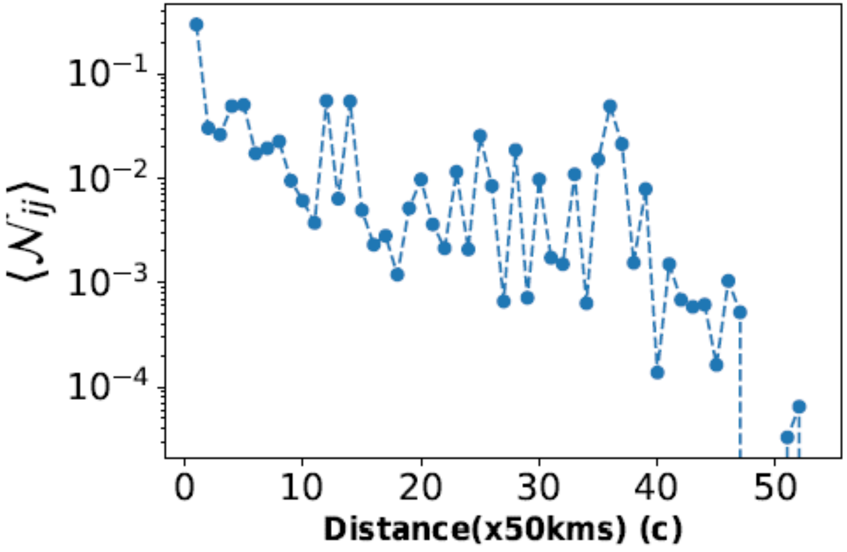
P_{ij} : common papers between i and j

w_i : papers by i

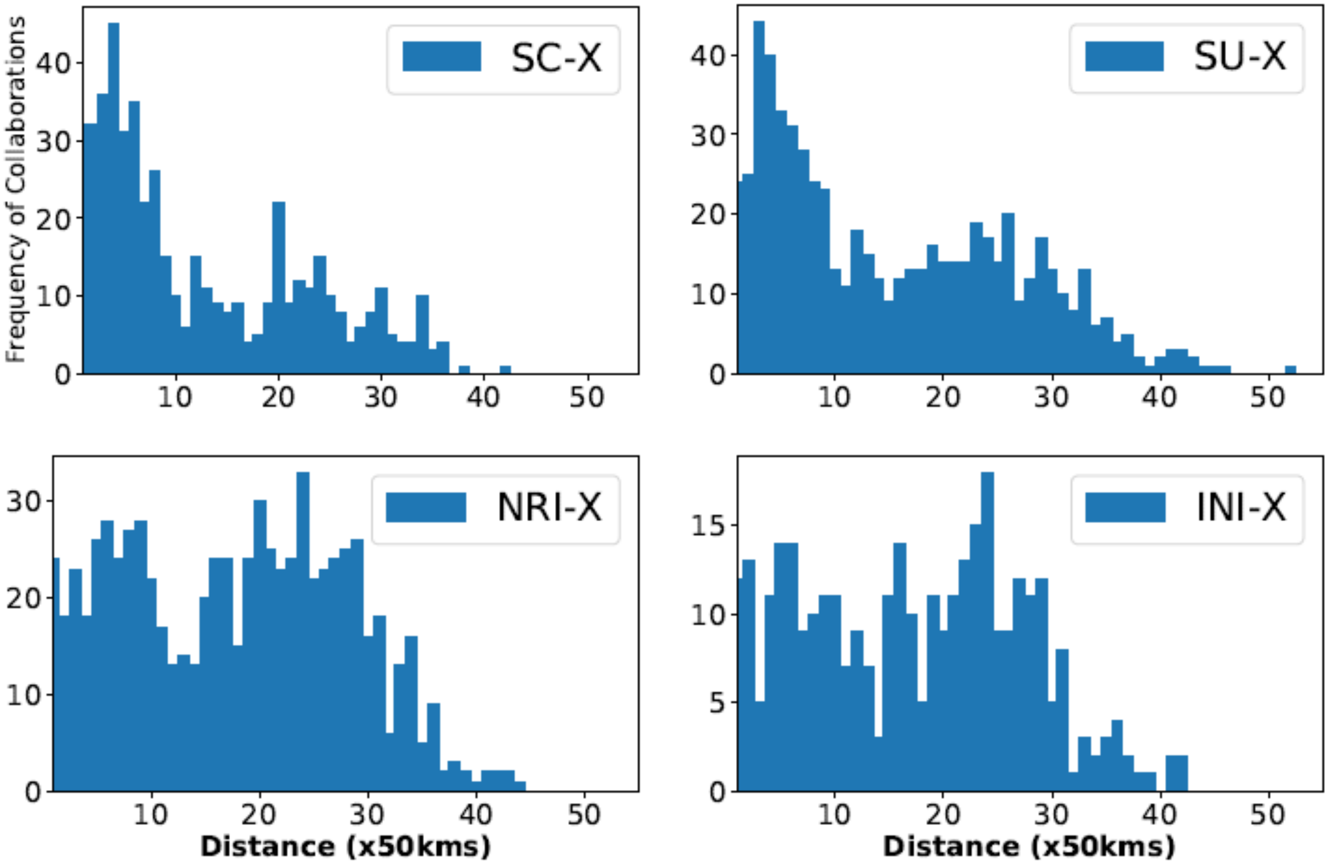
w_j : papers by j


$\mathcal{N}_{ij}$
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 52
Distance(X50kms) (a)





Notice that post 1995 the strength is less dependent on distance than pre 1995

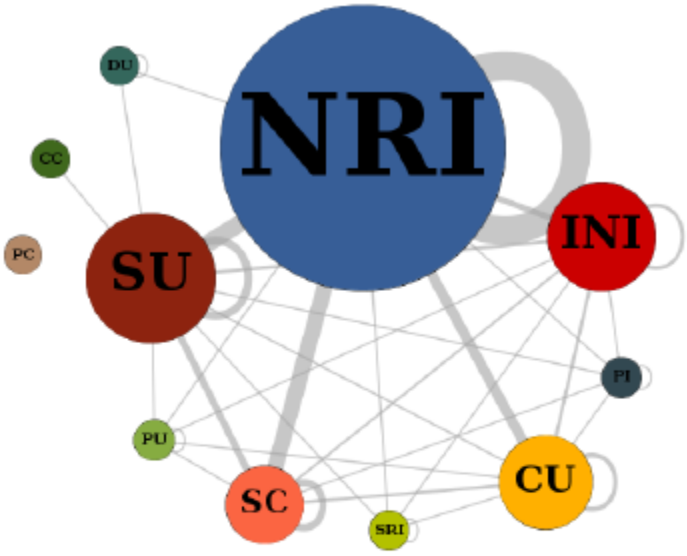

SC-X
SU-X
NRI-X
INI-X
Frequency of Collaborations
Distance (x50kms)
Distance (x50kms)

Does productivity promote collaboration ?

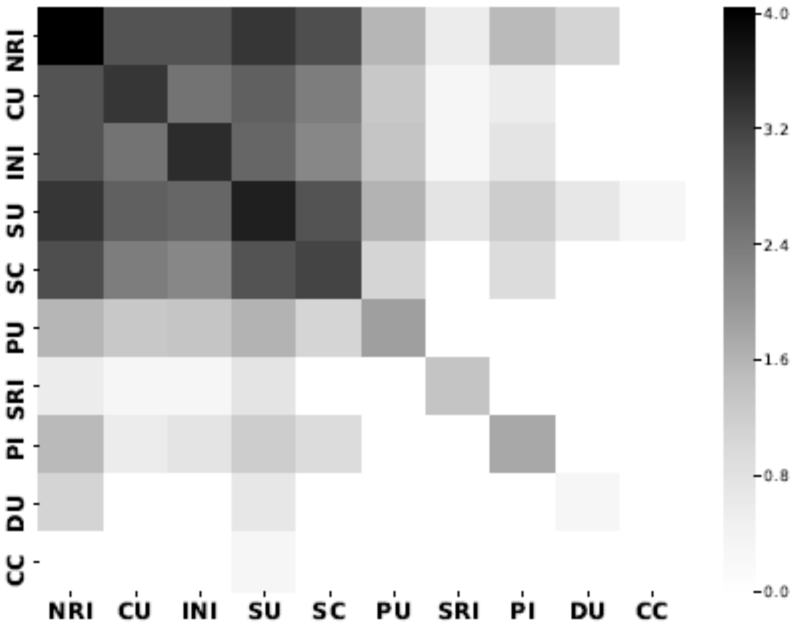
Table 3.1. Number of Papers from Different types of Institutes in the data set studied till 2013

	NRI	INI	CU	SU	SC	CC	DU	PU	PI	SRI
Papers	9292	2635	2083	3438	1482	1	9	85	57	25
Institutions	76	46	32	109	301	1	4	18	19	6
Papers per Institute	122.3	57.3	65.1	31.5	4.9	1	2.25	4.7	2.68	4.17

a



b



(a) Collaboration between Institutes aggregated by their category. Size represents productivity and edge weight represents strength in collaboration

(a) The Adjacency matrix where nodes are ordered in decreasing order of productivity.

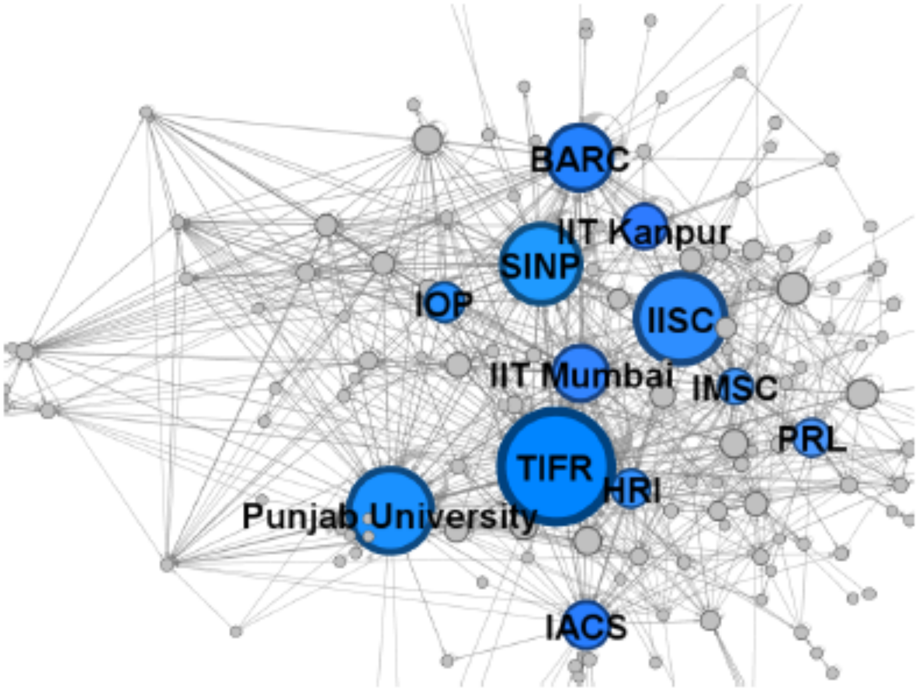
Knowledge flow

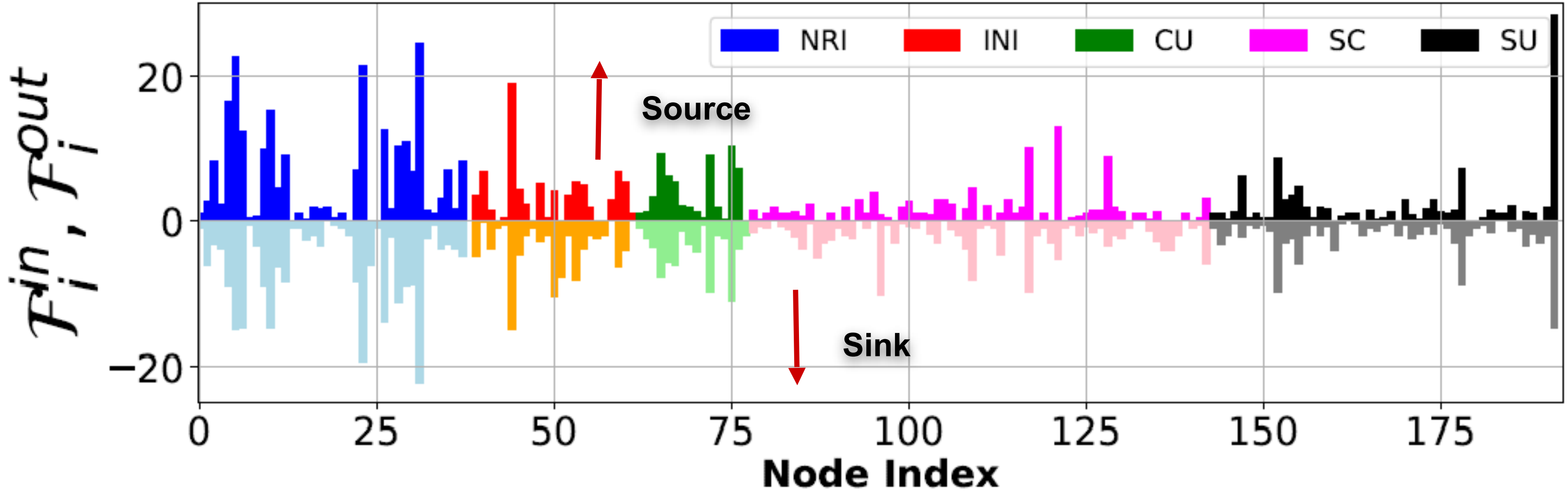
$$\mathscr{F}_i^{in} = -k_i^{out} \times \frac{W_i^{out}}{W_i^{in} + W_i^{out}}$$

$$\mathscr{F}_i^{out} = k_i^{in} \times \frac{W_i^{in}}{W_i^{in} + W_i^{out}}$$

$k_i^{out/in}$: out degree/ in degree

$W_i^{out/in}$: weighted out degree/ in degree





NRI
INI
CU
SC
SU
Source
Sink
20
0
−20
0
25
50
75
100
125
150
175
Node Index
$\mathcal{F}_i^{in}, \mathcal{F}_i^{out}$

Summary of Key Findings

- NRI's and INI's are dominant in institute collaboration networks in our dataset
- No strict evidence of power-law in collaboration vs geographic proximity. (It depends on the structure of the institution)
- Rich club - Highly productive institutes collaborate more within themselves (However this could be an artefact of the dataset)


One hour on this planet is 7 years on Earth
Great! I'll do my PhD here

Springer Link

Published: 24 July 2020

Evolution of interdependent co-authorship and citation networks

Chakresh Kumar Singh, Demival Vasques Filho, Shivakumar Jolad & Dion R. J. O'Neale

Scientometrics **125**, 385–404(2020) | Cite this article

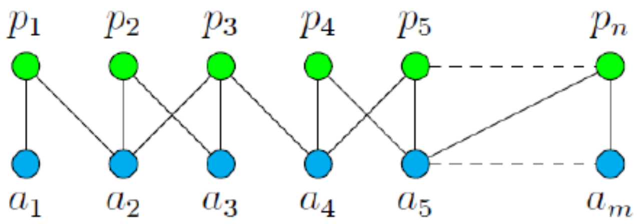
DOI : https://doi.org/10.1007/s11192-020-03616-0

Research questions

Interdependence between citation and co-authorship networks

- Effect of co-authorship distances
- Proportion of Self Citations
- Reference count
- Reciprocity in Citations
- Aging Effect

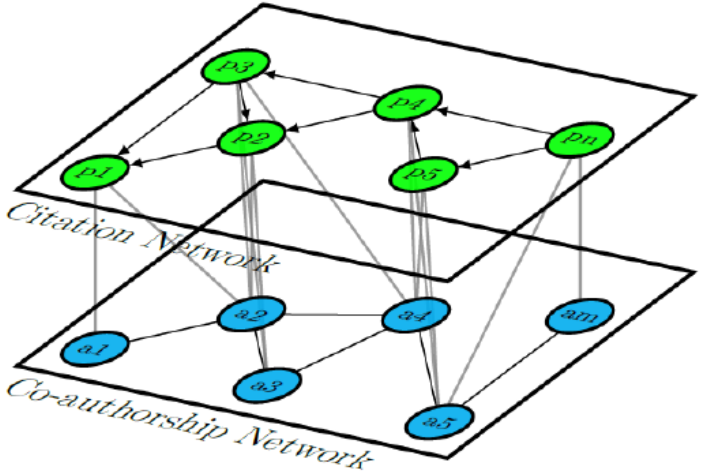
Networks



Bipartite Network of papers and authors

Adding citations

Multilayer Representation



Matrices

D

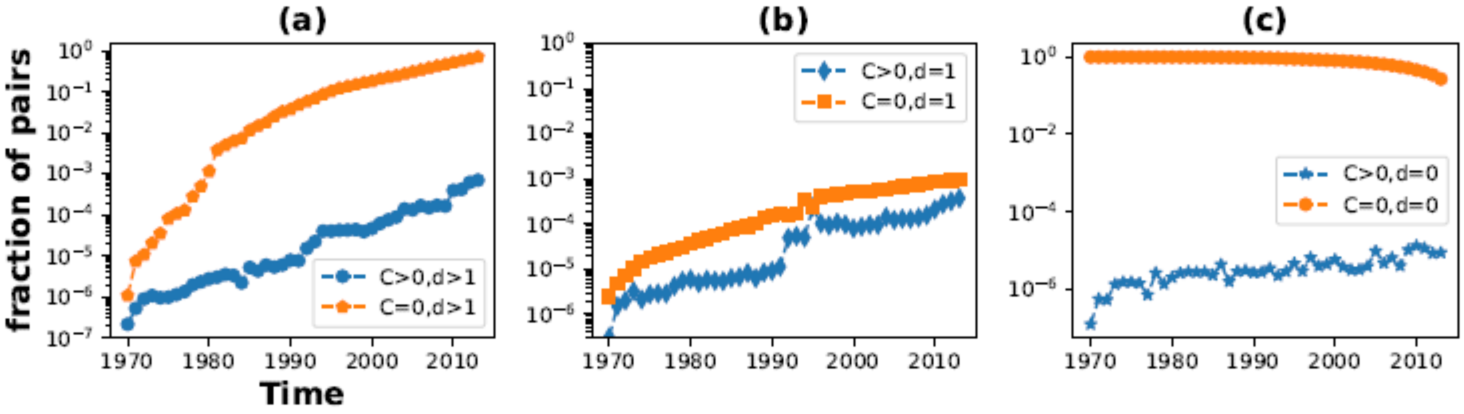
$$d_{ij}(t) = \begin{cases} d \text{ if there is a shortest path of length } d \text{ from } i \text{ to } j \\ 0 \text{ otherwise}, \end{cases}$$

C

$$c_{i \leftarrow j}(t) : \text{No.of times j cites i uptill time t}$$

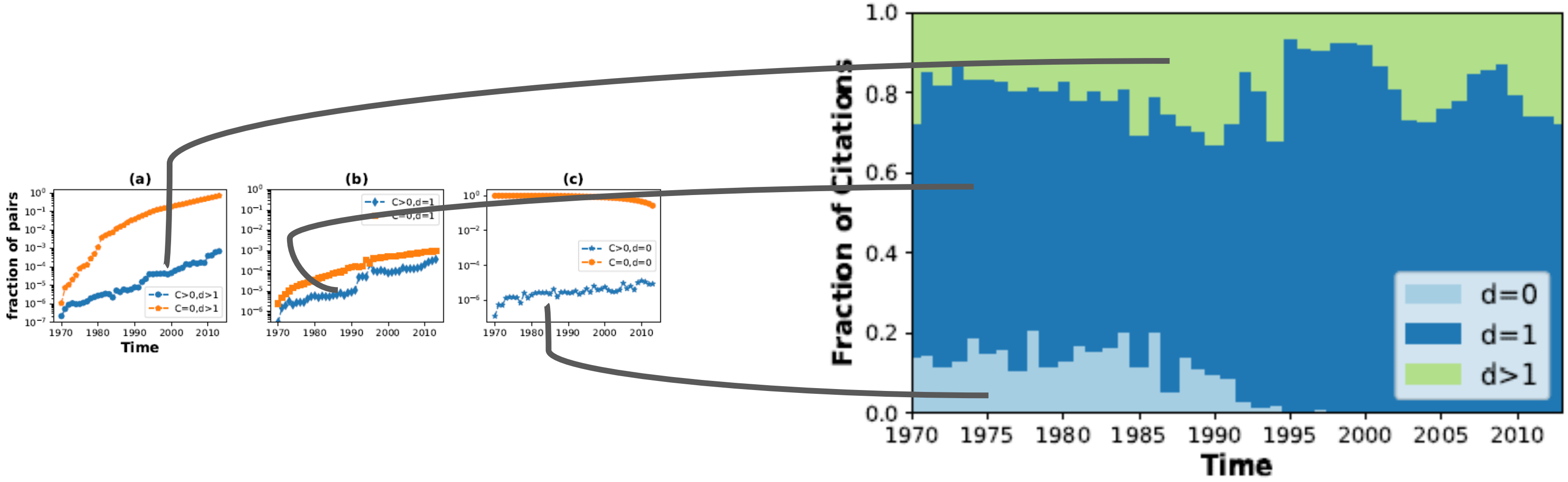
Note that by constructing the matrices D and C at every time step we can trace the history of all possible pairs in our network.

Citations at different co-authorship distances



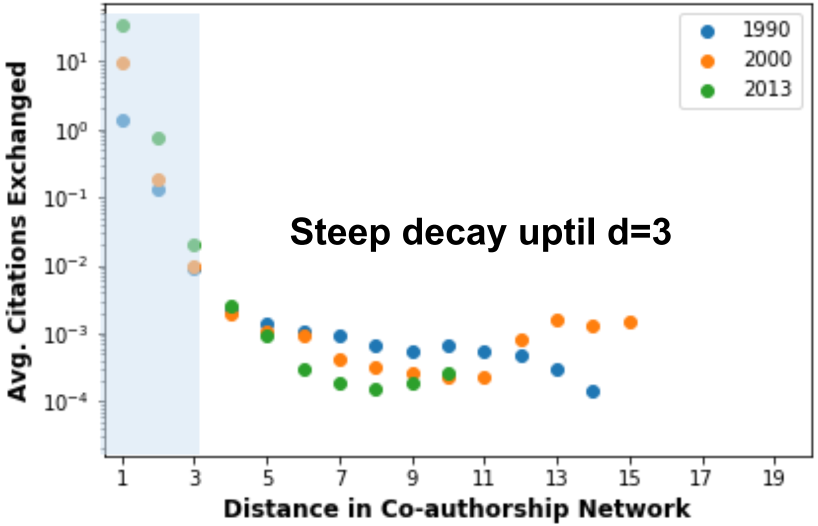
We are interested in Blue curves

Citations at different co-authorship distances



Most citations interactions are explained by your co-authors

Citations Exchanged fall rapidly with distance



Null Model

Note that here we restrict to a closed community i.e Indian Researchers

Popularity Generosity

$$P_{i \leftarrow j}(t) = \boxed{f_i(t)} \times \boxed{c^j(t^*)}$$

$$f_i^c(t) = \frac{\sum_j c_{ij}(t)}{\sum_i \sum_j c_{ij}(t)},$$

$$c^j(t^*) = \frac{\sum_i c_{ij}(t^*)}{\sum_i \sum_j c_{ij}(t^*)},$$

$$c_{ij}(t^*) = c_{ij}(t) - c_{ij}(t-1).$$

We construct a simple model for citations exchange between two authors <u>without</u> considering their relationship in the co-authorship network

Conditional probabilities

$$P(d=k|C^{>0}(t)) = \frac{P(C^{>0}(t)|d=k).P(d=k)}{P(C^{>0}(t))}$$

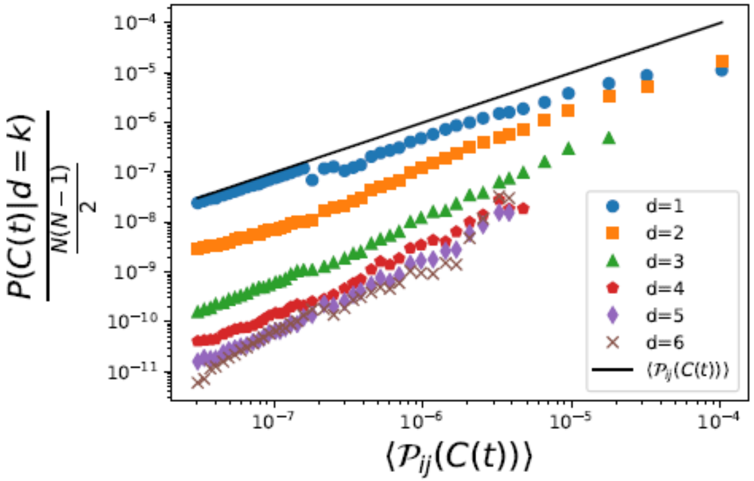
$$P(C^{>0}(t)) = \sum_{k=0}^{\infty} P(C^{>0}(t)|d=k) \times P(d=k)$$

$C^{>0}(t)$: Pairs with >0 citations exchanged

$C^{0}(t)$: Pairs with 0 citations exchanged

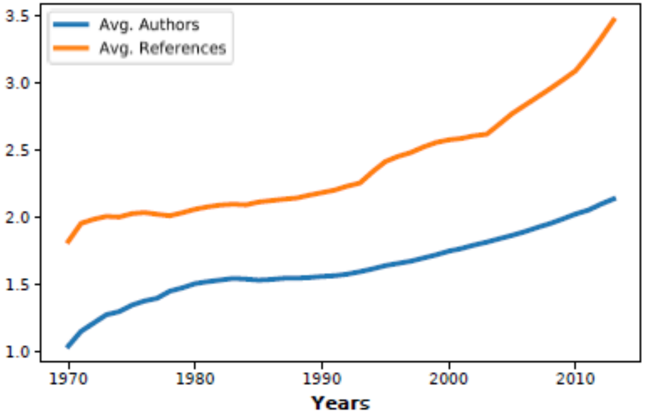
$$P(C^{>0}(t)|d=k) = \frac{P(d=k|C^{>0}(t)) \times P(C^{>0}(t))}{P(d=k|C^{>0}(t))P(C^{>0}(t)) + P(d=k|C^{0}(t))P(C^{0}(t))}$$

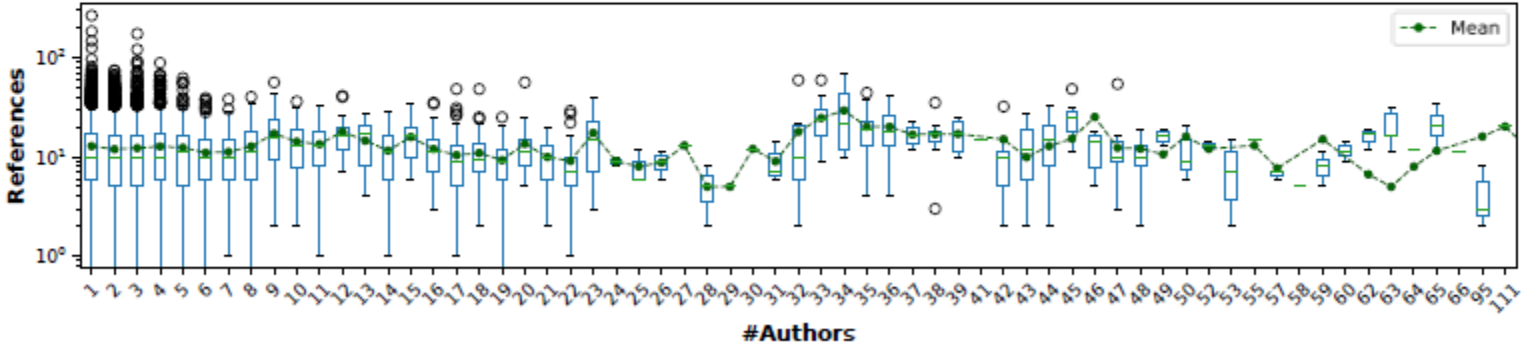
- **Simple model is not sufficient to explain the citations exchanged for higher distances**
- **Is this an expected behavior?**



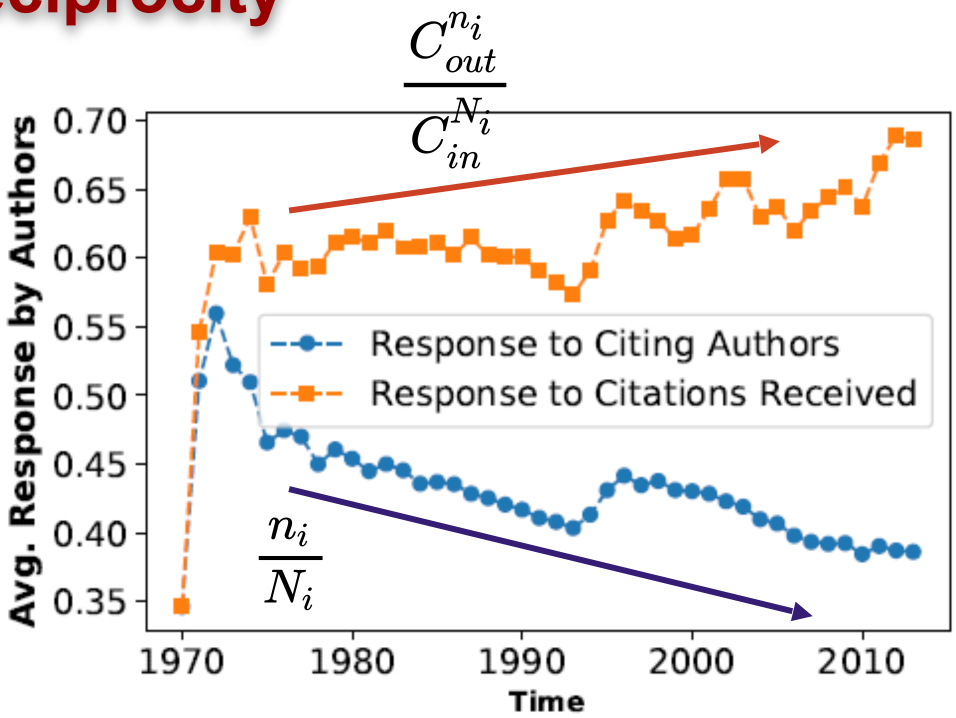
References vs. Authors

- Avg. references per paper and Avg. authors per paper increase overtime
- No evident correlation between the two





Reciprocity



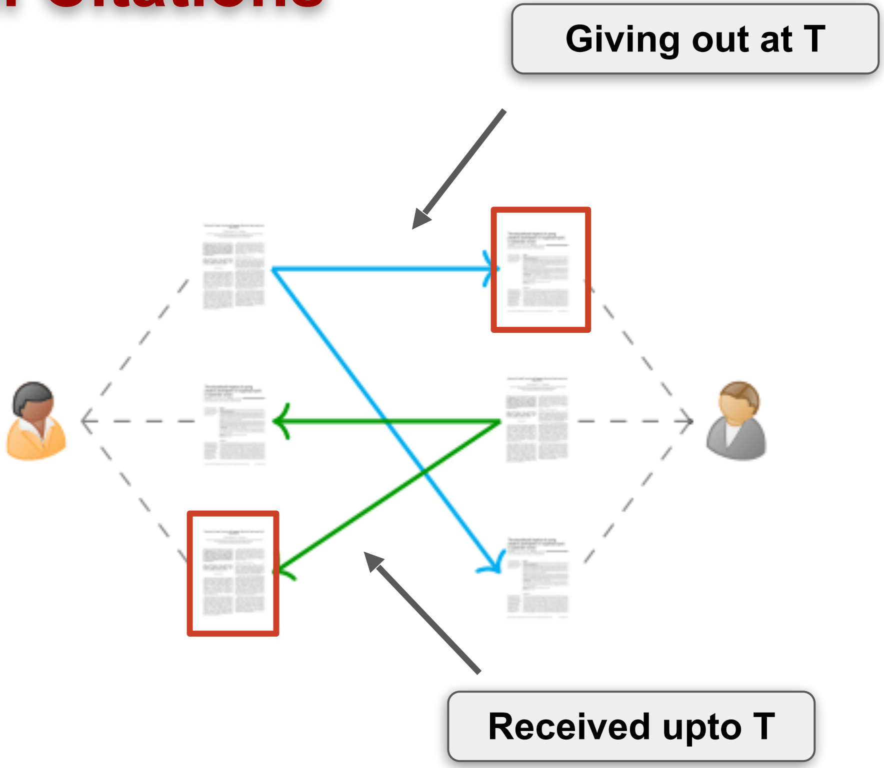
N_i : authors that cite i

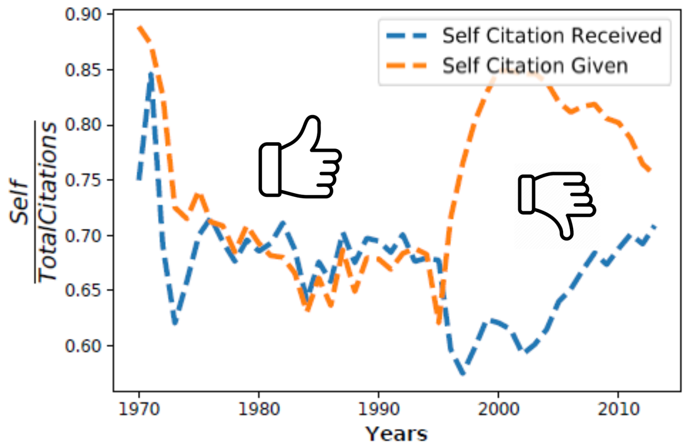
n_i : subset of N_i whome i cites back

$C_{out}^{n_i}$: citations given out to n_i

$C_{in}^{N_i}$: citations received from N_i

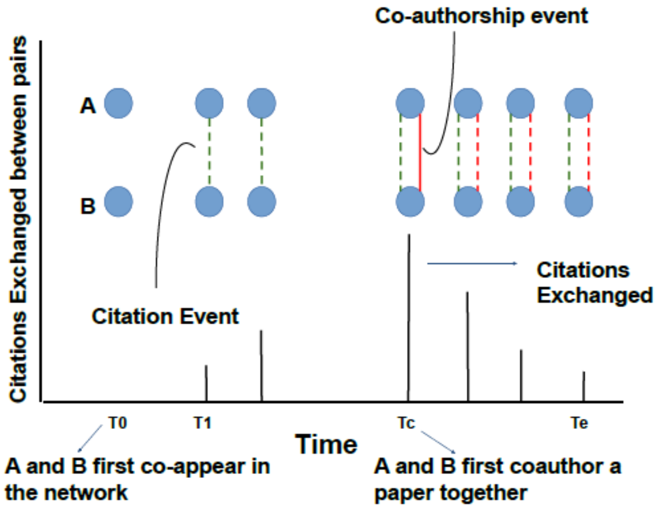
Self Citations

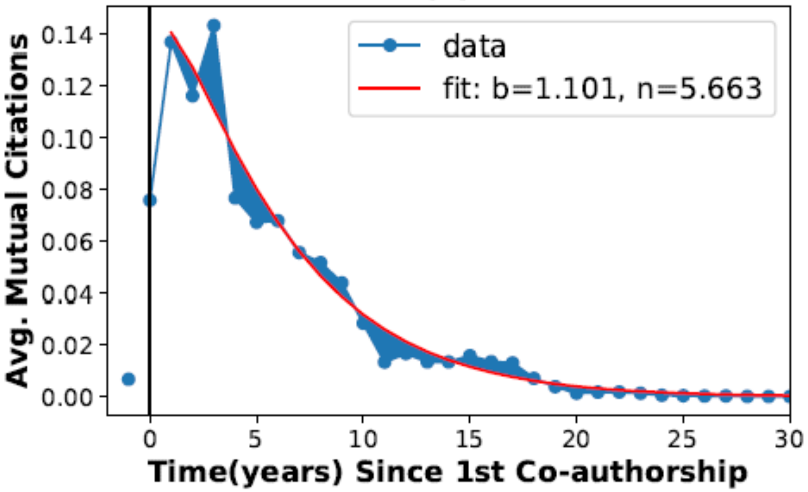




Aging in collaboration

$$f(t) = \frac{b}{n}\left(\frac{t}{n}\right)^{b-1} e^{-\left(\frac{t}{n}\right)^{b}}$$





Summary of Key Findings

- In a closed community most citations are explained by your close collaborators.
- Distant collaborators (d>3) have similar effect
- Reciprocity observes a decaying trend (artifact of aging in citations)
- Strength of collaboration weakens over time.
- <u>Evolution of Collaboration networks are better understood in an interdependent setting</u>

Impact of this work

1. Aid in assessing the growth of science in a country such as India and can help the policymakers to frame policies for the advancement of science and technology.
2. From a historic perspective, such types of extensive and focussed longitudinal studies can supplement other works that look into how science and society affect each other's evolution.
3. The methodology implemented in this thesis is independent of the type of networks used, therefore, the same could be used to understand collaborations in different form of networks.
4. This would help in comparing trends between various forms of correlated temporal complex networks and contribute in understanding the complexity of real-world systems in particular to the science of science

Future Perspectives

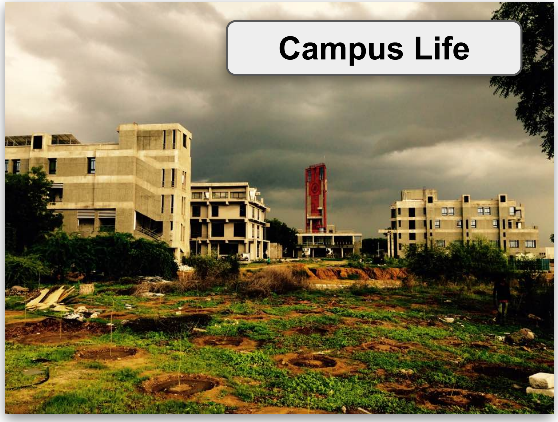
- Expanding the analyses to very large datasets which we could not perform due to limited access.

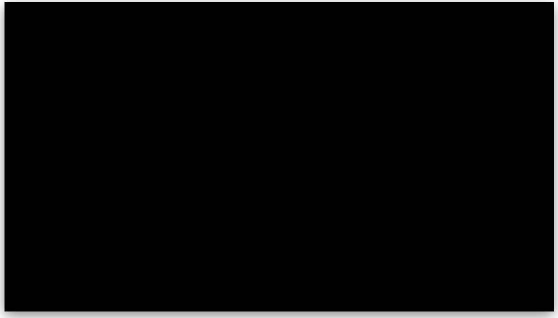
- A generic modeling framework to understand collaboration patterns at different scales.

Acknowledgement - I

- Prof. Shivakumar Jolad for his freedom, guidance and support.
- Prof. Krishna Kanti Dey for guidance and support at most crucial phase.
- Prof. Anand Sengupta and Prof. Anirban Dasgupta for their valuable insights and support.
- Prof. Dion O’Neale and Demival V. Filho for their support and collaboration
- Special thanks to Prof. Sudipta Sarkar, Prof. Baradhwaj Coleppa and Prof. Vinod Chandra for their motivation and financial support for travel.
- IITGN - for all the support and financial help without which things would have been tough

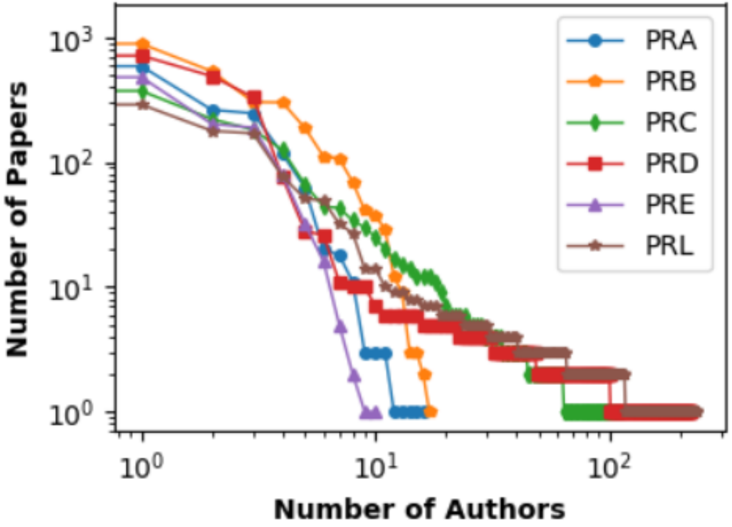
Acknowledgement - II

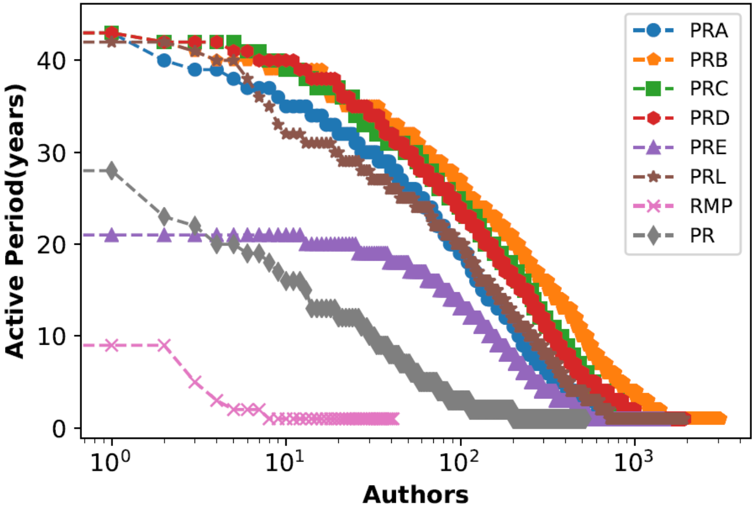




Dept. Dinners

Backup Slides





Low cut-off for PRA,PRB and PRE

RMP are mostly Review Papers

Diversity Calculation

To define the diversity of a paper and/or an author we have used Weitzman diversity [162, 163, 164] method. Weitzman diversity W_D measure for a set of species measures the diversity or dissimilarity in the set given there is a unique distance metric defined between its elements. The distance metric should follow :

$$d(i,j) \geq 0 \tag{6.5}$$

$$d(i,i) = 0 \tag{6.6}$$

$$d(i,j) = d(j,i) \tag{6.7}$$

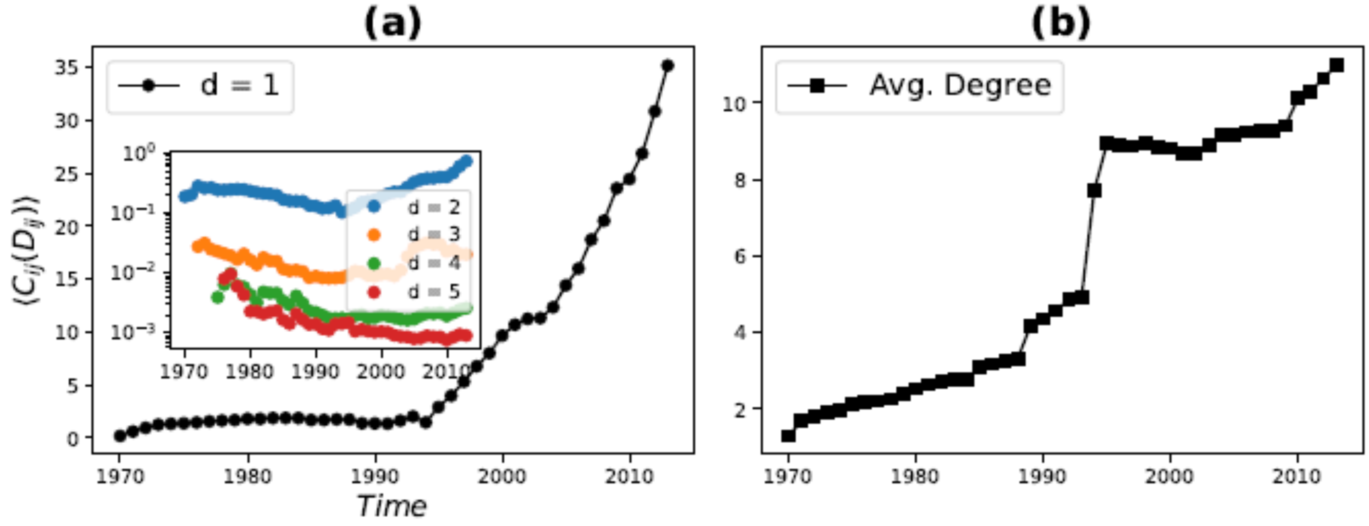
The definition of the distance metric is not directly relevant for the measure of the diversity as long it meaningfully measures the difference between the elements. Let S be a set of species and Q a subset of S. Let j be any element belonging to S but not to Q i.e. $j \in S \backslash Q$. Then

$$d(j,Q) \equiv min_{i \in Q} d(i,j) \tag{6.8}$$

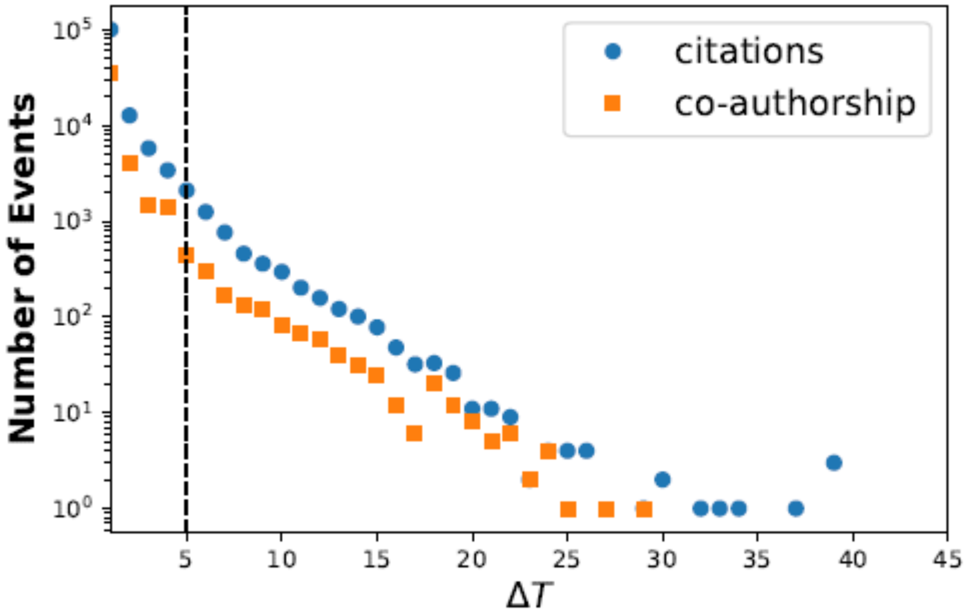
We begin with a set of PACS S of a paper and an empty set X. Recursively we pick an element u_i from S and place it in X. Then the Weitzman diversity $W_D(S)$ for set S is defined as follows:

$$W_D(S) = \sum_{i=1}^{N} d(u_i, X) \tag{6.9}$$

where $d(u_i, X)$ is calculated using Eq. 6.8 and is the change in diversity due to addition of new element in X. This is continued till all elements in S are added to X.

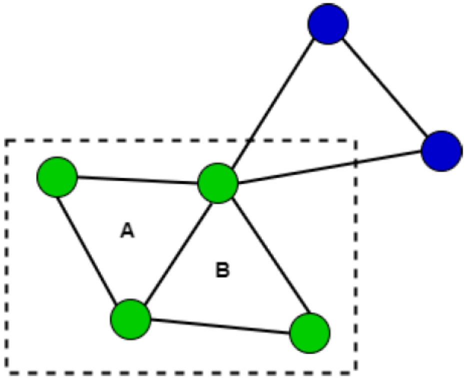


Average citations exchanged with distances

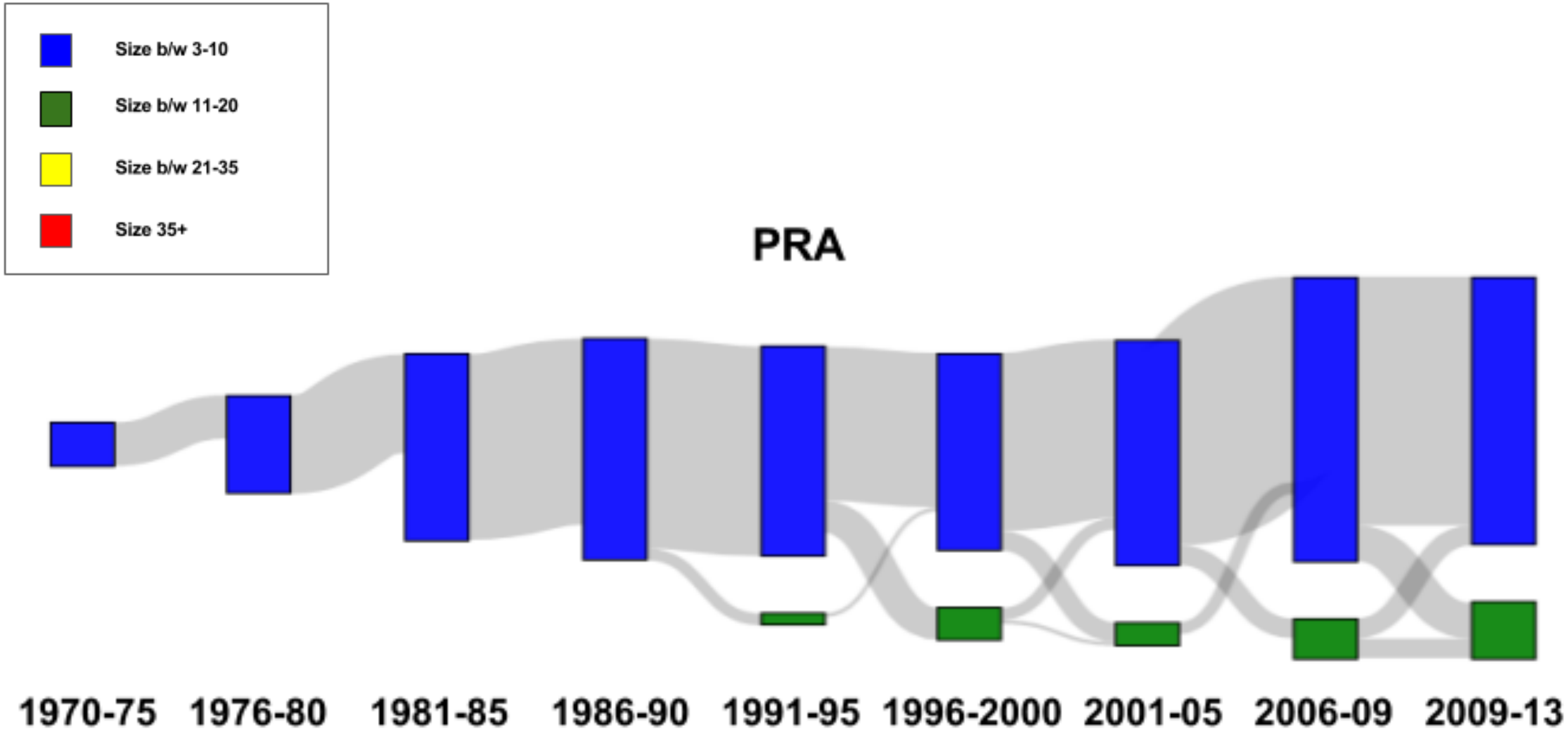


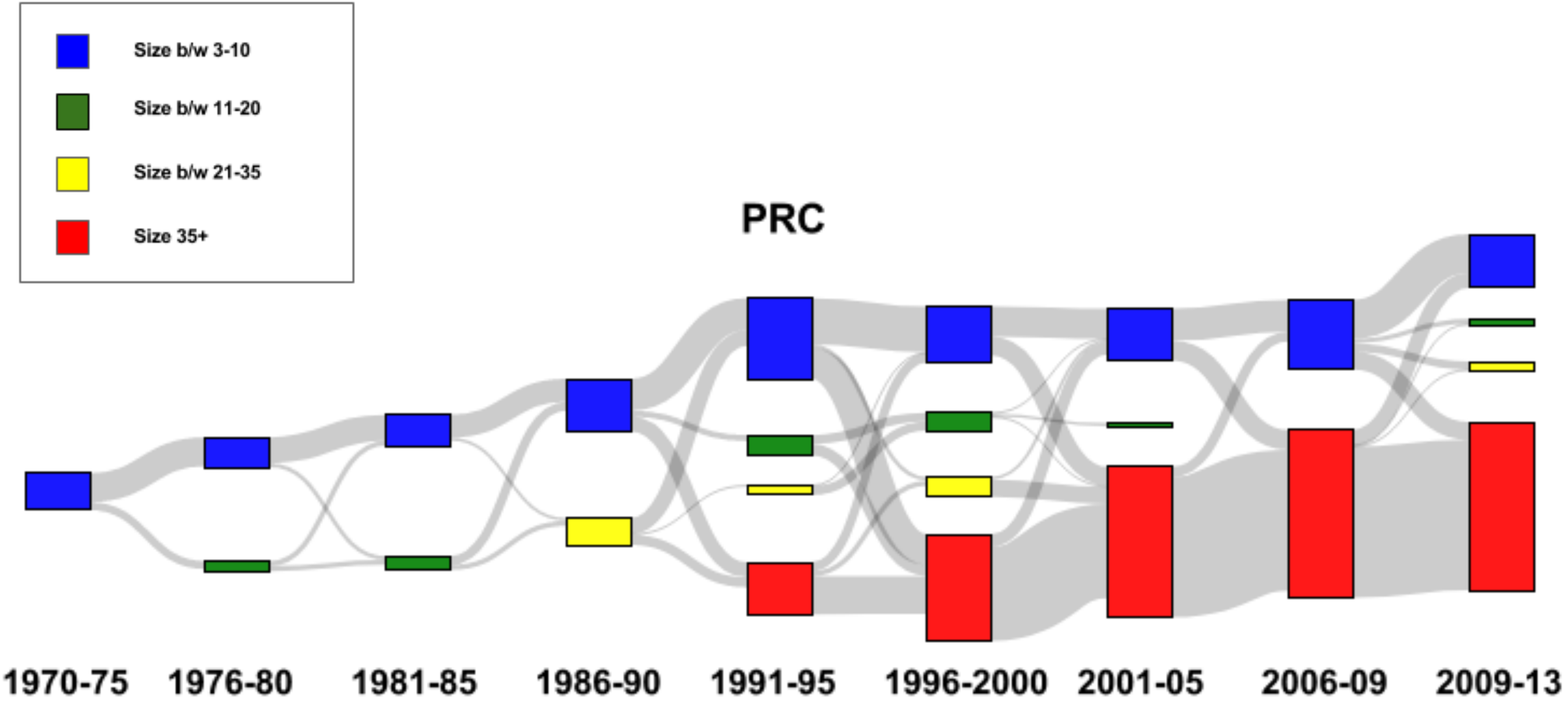
Number of Events
10^5
10^4
10^3
10^2
10^1
10^0
citations
co-authorship
5
10
15
20
25
30
35
40
45
ΔT

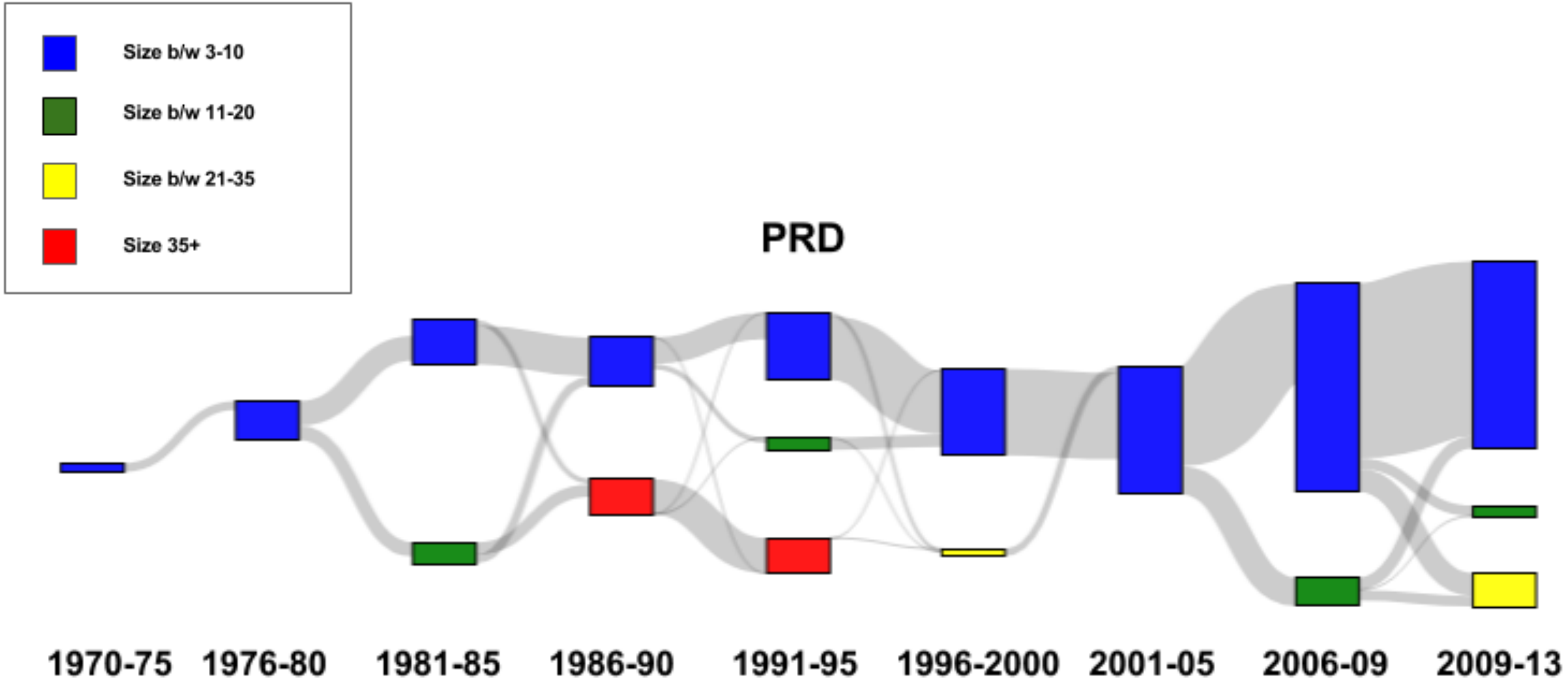
CPM

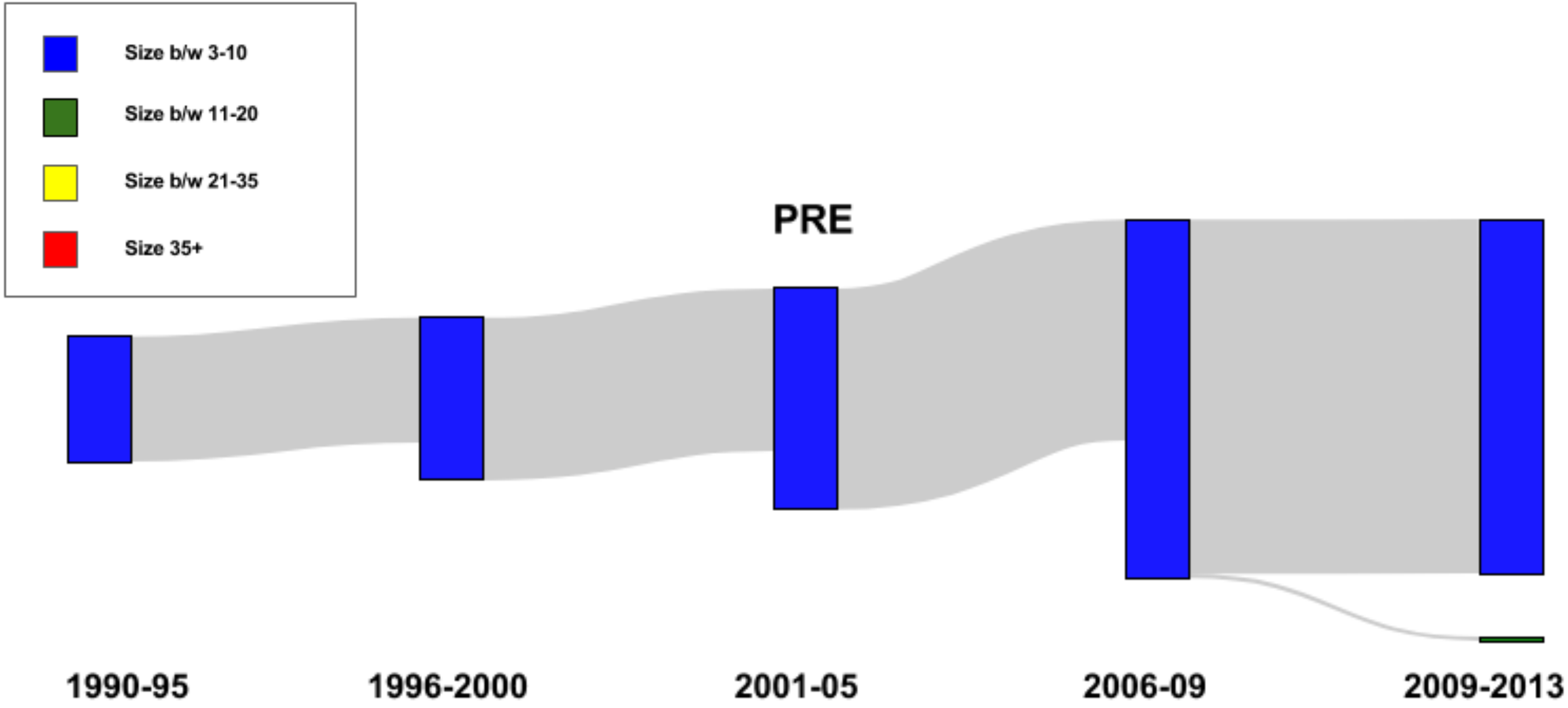


- Detect cliques of sizes $\geq k$, where k is the minimum size of the community. k cannot be lower than 3.
- Create a matrix O^c, where O^c_{ij} is the number of vertices shared by clique i and j.
- Choose a value of k which would determine the resolution for detecting communities.
- Set the diagonal elements that are $< k$ of O^c to zero.
- Set the off-diagonal elements that are $< k-1$ of O^c to zero.
- Now set all non-zeros in O^c to 1. The resultant O^c is the clique-clique overlap matrix.
- Run clustering on this matrix to get the community structure.

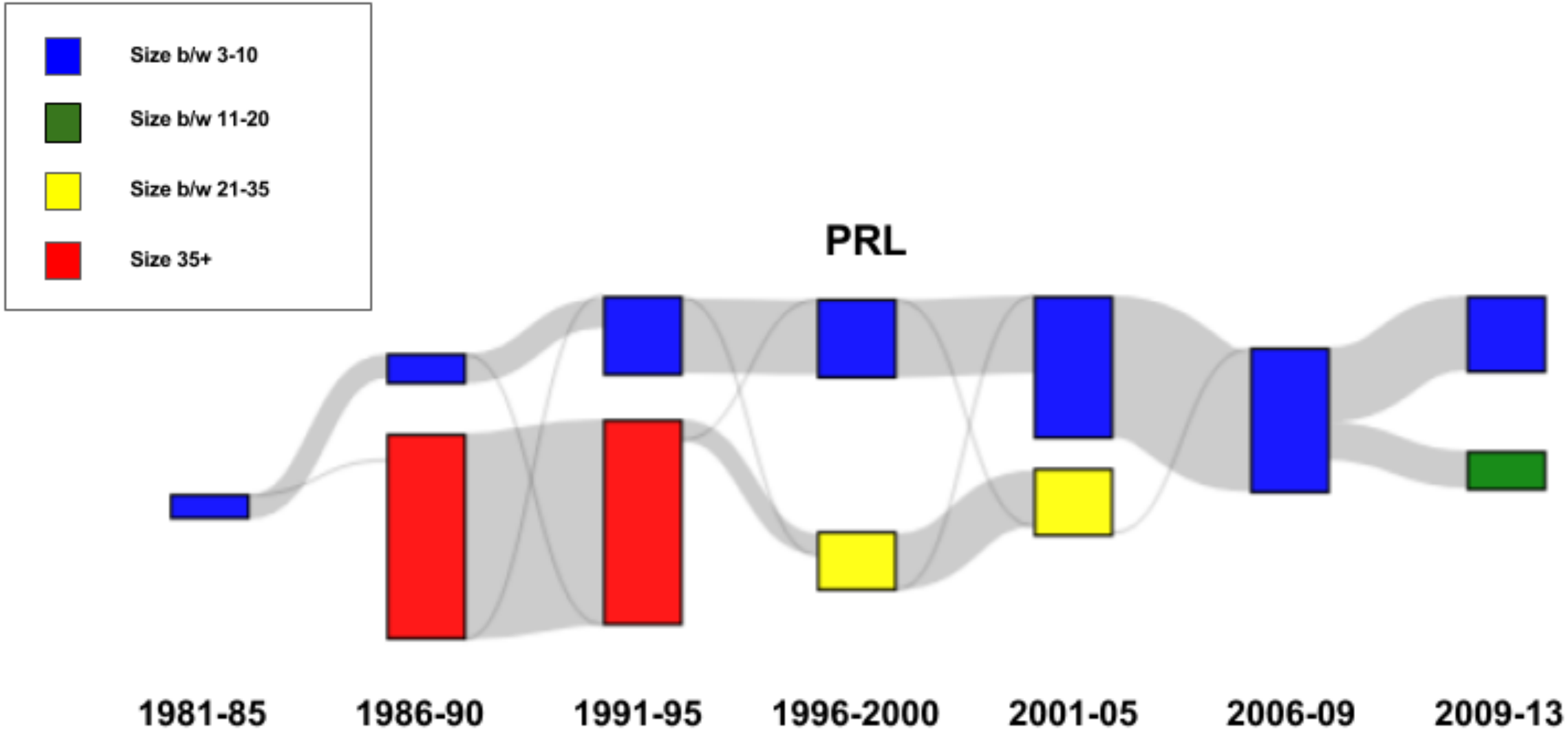

Size b/w 3-10
Size b/w 11-20
Size b/w 21-35
Size 35+
PRA
1970-75
1976-80
1981-85
1986-90
1991-95
1996-2000
2001-05
2006-09
2009-13

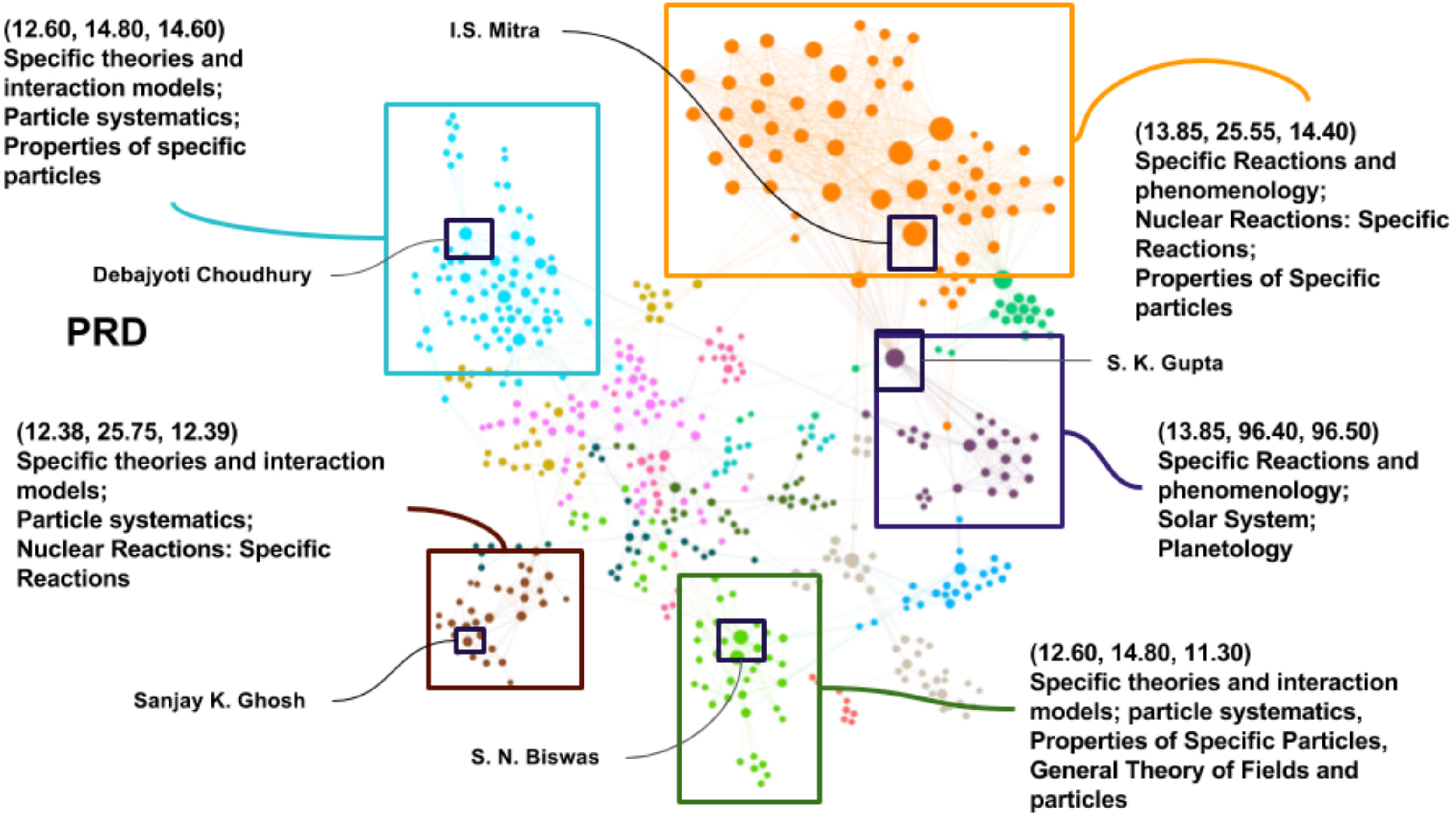

Size b/w 3-10
Size b/w 11-20
Size b/w 21-35
Size 35+
PRC
1970-75
1976-80
1981-85
1986-90
1991-95
1996-2000
2001-05
2006-09
2009-13


Size b/w 3-10
Size b/w 11-20
Size b/w 21-35
Size 35+
PRD
1970-75
1976-80
1981-85
1986-90
1991-95
1996-2000
2001-05
2006-09
2009-13



Size b/w 3-10
Size b/w 11-20
Size b/w 21-35
Size 35+
PRE
1990-95
1996-2000
2001-05
2006-09
2009-2013


Size b/w 3-10
Size b/w 11-20
Size b/w 21-35
Size 35+
PRL
1981-85
1986-90
1991-95
1996-2000
2001-05
2006-09
2009-13



(12.60, 14.80, 14.60)
Specific theories and interaction models;
Particle systematics;
Properties of specific particles
I.S. Mitra
Debajyoti Choudhury
PRD
(13.85, 25.55, 14.40)
Specific Reactions and phenomenology;
Nuclear Reactions: Specific Reactions;
Properties of Specific particles
S. K. Gupta
(12.38, 25.75, 12.39)
Specific theories and interaction models;
Particle systematics;
Nuclear Reactions: Specific Reactions
(13.85, 96.40, 96.50)
Specific Reactions and phenomenology;
Solar System;
Planetology
Sanjay K. Ghosh
S. N. Biswas
(12.60, 14.80, 11.30)
Specific theories and interaction models; particle systematics,
Properties of Specific Particles,
General Theory of Fields and particles

Comm. Matching

[Link to paper]

To overcome this difficulty, we refine the identification of communities as shown in Fig.1f in the main text and in Fig.8. For each consecutive time steps t and $t+1$ we construct a joint graph consisting of the union of links from the corresponding two networks, and extract the CPM community structure of this joint network. When new links are introduced in a network, the CPM communities may remain unchanged, they may grow, or a group of CPM communities may become joined into a single community, however no CPM community may decay by losing members. From this it follows that if we merge two networks, any CPM community in any of the original networks will be contained in exactly one community in the joined network. Let us denote the set of communities from t by $\mathbf{D}$, the set of communities from $t+1$ by $\mathbf{E}$, and the set of communities from the joint network by $\mathbf{V}$. For any community $D_i \in \mathbf{D}$ or $E_j \in \mathbf{E}$ we can find exactly one community $V_k \in \mathbf{V}$ containing it. A very important point is that when checking whether D_i or E_j is contained or not in V_k, we compare the *links* in the corresponding communities. The CPM permits a community to contain even all nodes of another community, (*e.g.* if

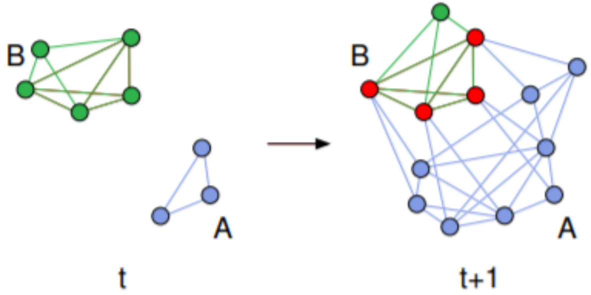


Figure 7: A situation where simple relative node overlap between the communities is not enough to decide the matching between t and $t+1$. (The k-clique size is $k=3$.)

in Fig.7. the community B would not gain a new member, then all nodes in B_{t+1} would be contained in A_{t+1} as well), therefore distinction between communities has to be made based on links rather then members. (In the studied systems an example where a smaller community can be formed “on top” of a larger community in this way could be a small group of friends working for the same large company, or a small group of scientists collaborating in an additional field of interest as well beside the main research area of their larger community).

Turning back to our matching algorithm: When matching the communities in $\mathbf{D}$ and in $\mathbf{E}$, first for every community $V_k \in \mathbf{V}$ in the joint system we extract the list of communities $D_i^k \in \mathbf{D}$ and $E_j^k \in \mathbf{E}$ that are contained in V_k (this means $D_i^k \subseteq V_k$ and $E_j^k \subseteq V_k$). (Note that either of the lists may be empty). Then the relative overlap between every possible (D_i^k, E_j^k) pairs can be obtained as

$$C_{ij}^k = \frac{\left|D_i^k \cap E_j^k\right|}{\left|D_i^k \cup E_j^k\right|}, \tag{4}$$

and we match the pairs of communities in descending order of their relative overlap. As an illustration of the above process, in Fig.8 we show three simple scenarios occurring in the community evolution of the phone-call network. In Fig.8a both lists D_i^k and E_j^k consist of only a single community, therefore these can be matched right away. However, in Fig.8b the D_i^k list contains two elements, let us denote the light blue community of size $s = 6$ by D_1^k and the dark blue community consisting of nine nodes by D_2^k. The corresponding E_j^k list contains a single community E_1^k having 15 members. The relative overlaps between the communities are given as $C_{1,1}^k = 2/5$ and $C_{2,1}^k = 3/5$. Since the $C_{2,1}^k$ relative overlap of the yellow E_1^k community with the dark blue D_2^k community is larger than the $C_{1,1}^k$ relative overlap with the light blue D_1^k, we assign E_1^k to D_2^k. As a consequence the light blue D_1^k community comes to the end of its life at t, and it is swallowed by D_2^k. The opposite process is shown in Fig.8c: in this case the D_i^k list consists of a single community D_1^k of size $s = 15$, whereas the E_j^k list has two elements, the yellow community labelled E_1^k with six members, and the orange community labelled E_2^k containing ten nodes. The relative overlaps are $C_{1,1}^k = 2/5$ and $C_{1,2}^k = 2/3$, therefore the D_1^k is matched to E_2^k, and E_1^k is treated as a new born community. In general, whenever the community V_k contains more communities from $\mathbf{D}$ than from $\mathbf{E}$, the communities D_i^k left with no counterpart from E_j^k finish their life's at t, and when V_k contains more communities from $\mathbf{E}$ than from $\mathbf{D}$, the communities E_j^k left with no counterpart from D_i^k are considered as new born communities.