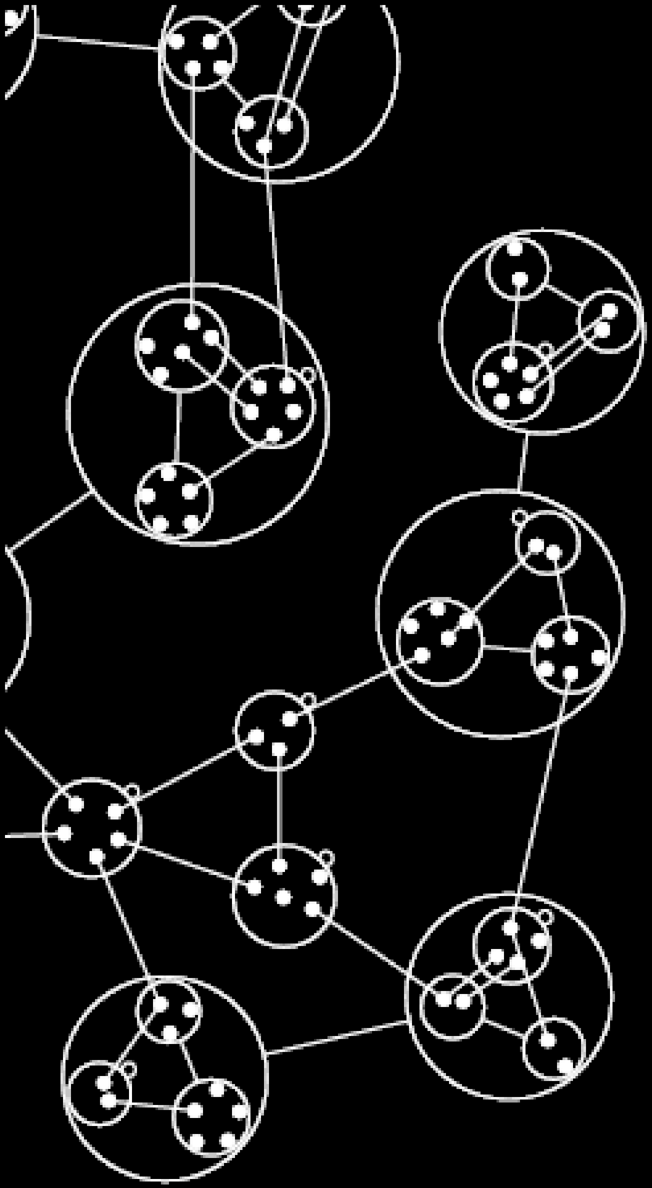


How to become a Group Leader?

Modeling Author Types based on Graph Mining

George Tsatsaronis, Iraklis Varlamis, Sunna Torge,
Matthias Reimann, Kjetil Nørvåg,
Michael Schroeder, Matthias Zschunke

Online Publication Data

Bibliographic Databases

Online Publication Data

Bibliographic Databases

dblp.uni-trier.de
Computer Science Bibliography
PubMed

Data mining

→

Social network analysis

Online Publication Data

Bibliographic Databases

Data mining

→

Social network analysis

Metadata based services

Arnetminer
Google scholar
Microsoft
Beta
Academic Search
CiteSeer
x
beta

Online Publication Data

Metadata based services

Arnetminer

Google scholar

Microsoft Academic Search Beta

CiteSeerx beta

Provide:

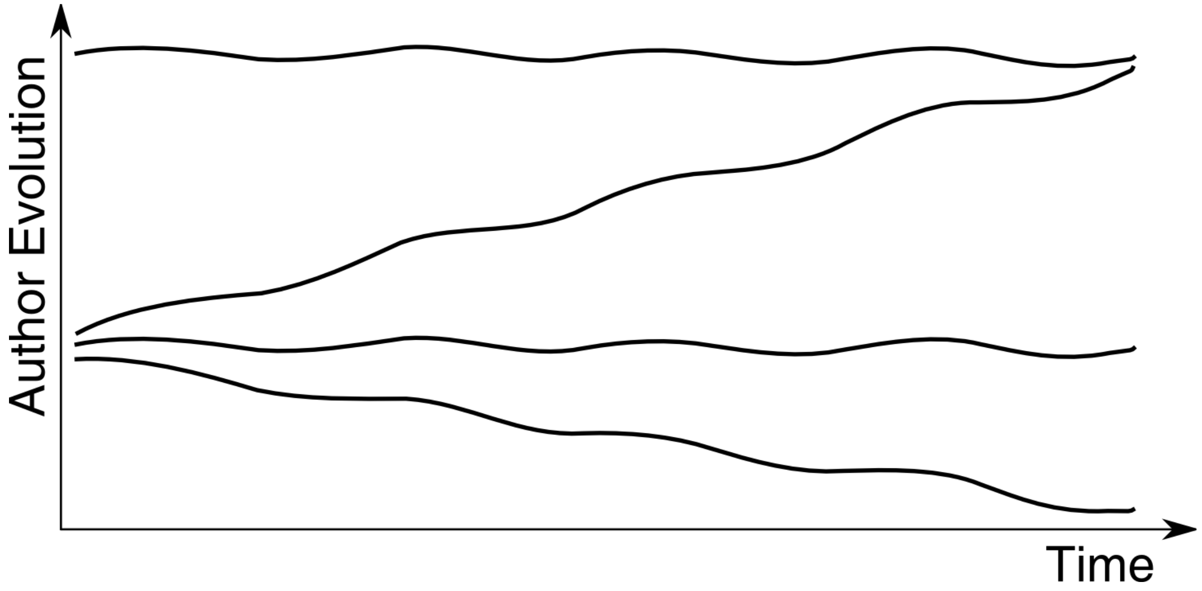
- Overall achievements
- List of publications
- List and network of co-authors
- H-Index

NOT:

- Evolution over time of
  - Publication list
  - Co-author network

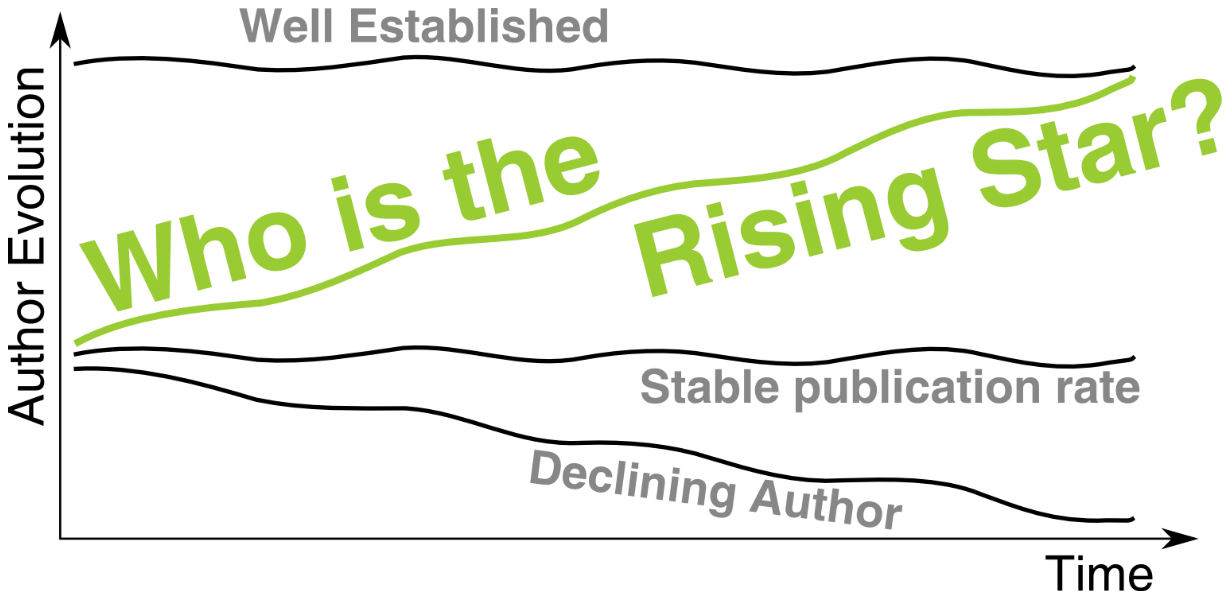
Objectives

1) Uncover **evolution** of publication records over time



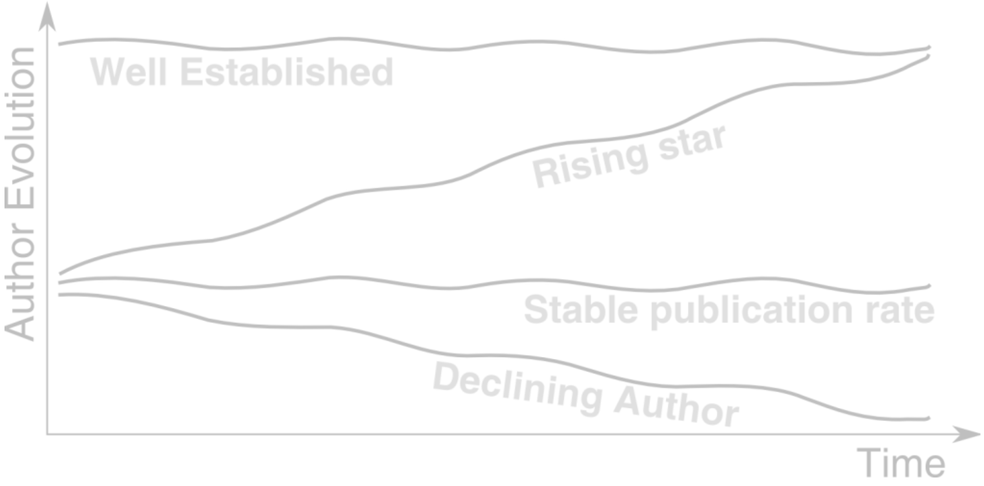
Objectives

1) Uncover **evolution** of publication records over time
2) Model basic **author types**



Methodology

1) Application of **Power Graph Analysis** to co-authorship graphs
2) Definition of **author features** based on power graph structure, number & impact of publications
3) Computation of **evolution indices (EI)**
4) **Clustering** based on EI
5) Employ EI of clusters as descriptors of **author types**



Methodology

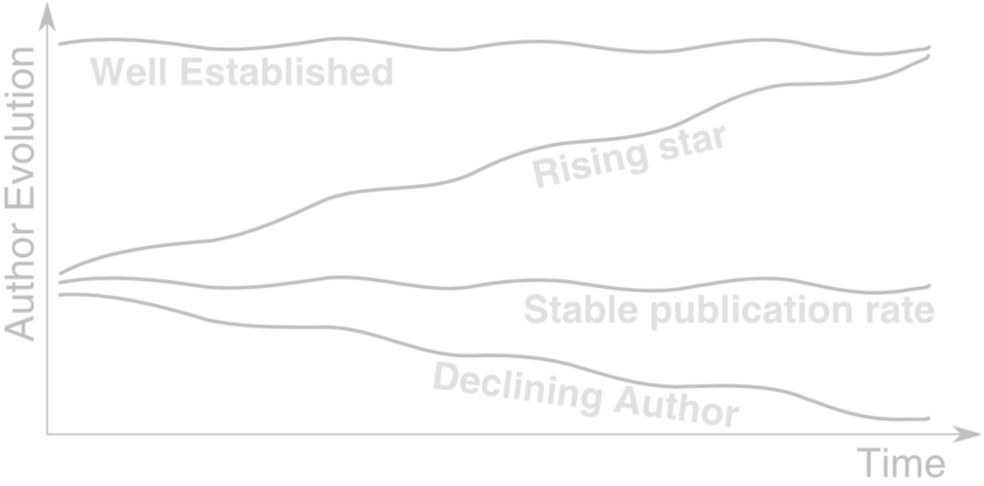
1) Application of **Power Graph Analysis** to co-authorship graphs

2) Definition of **author features** based on power graph structure, number & impact of publications

3) Computation of **evolution indices (EI)**

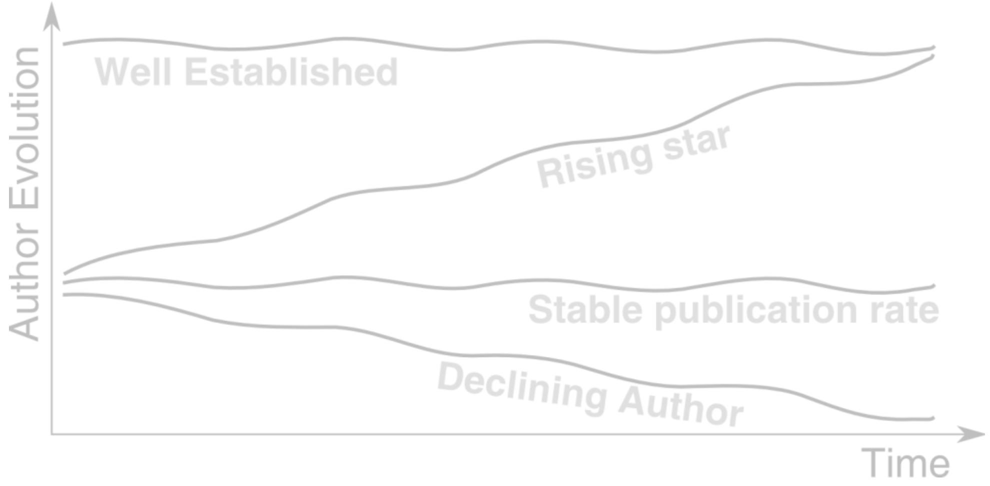
4) **Clustering** based on EI

5) Employ EI of clusters as descriptors of **author types**



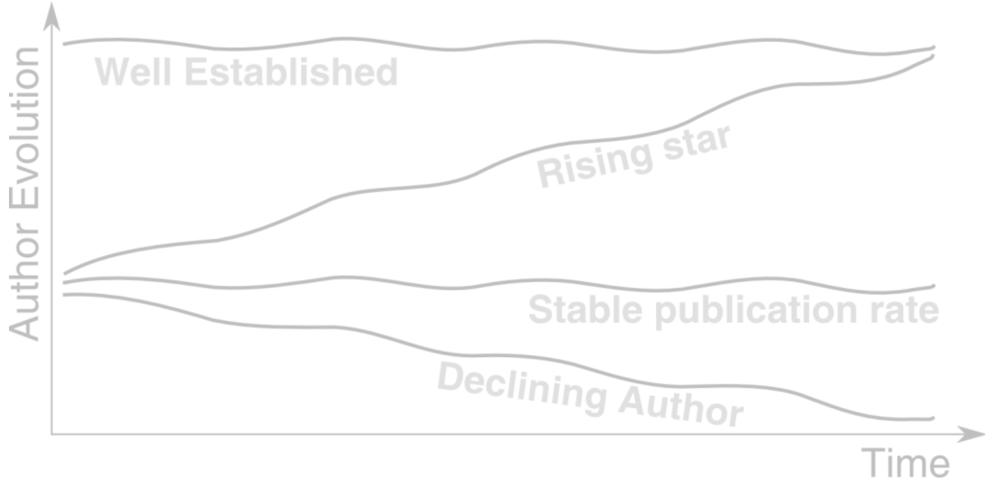
Methodology

1) Application of **Power Graph Analysis** to co-authorship graphs
2) Definition of **author features** based on power graph structure, number & impact of publications
3) Computation of **evolution indices (EI)**
4) **Clustering** based on EI
5) Employ EI of clusters as descriptors of **author types**



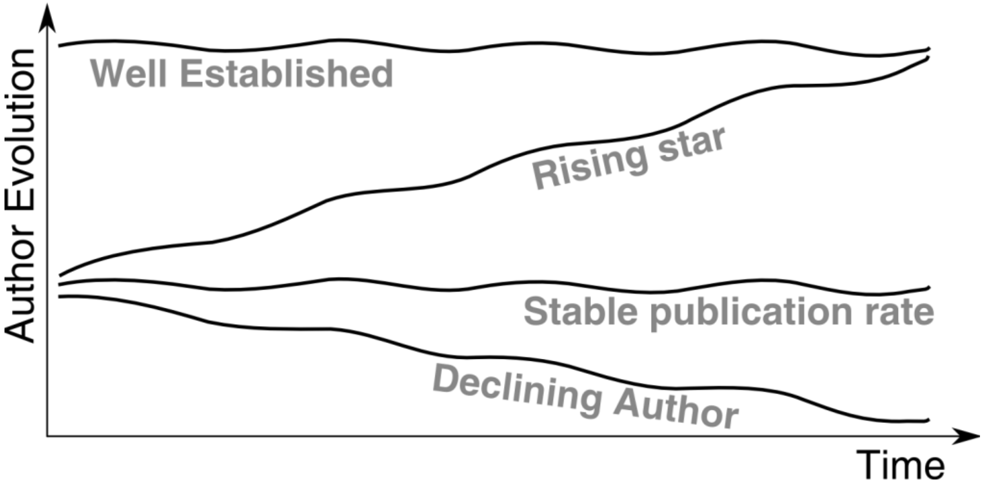
Methodology

1) Application of **Power Graph Analysis** to co-authorship graphs
2) Definition of **author features** based on power graph structure, number & impact of publications
3) Computation of **evolution indices (EI)**
4) **Clustering** based on EI
5) Employ EI of clusters as descriptors of **author types**



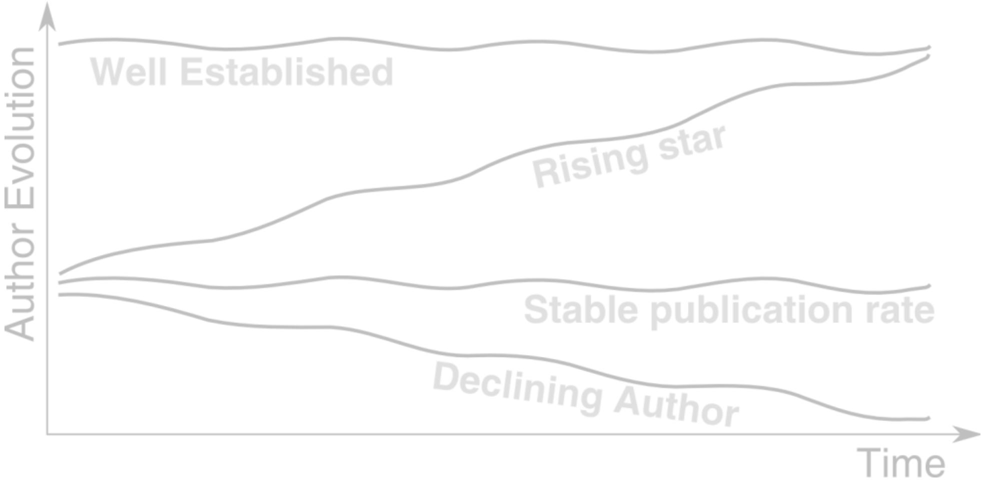
Methodology

1) Application of **Power Graph Analysis** to co-authorship graphs
2) Definition of **author features** based on power graph structure, number & impact of publications
3) Computation of **evolution indices (EI)**
4) **Clustering** based on EI
5) Employ EI of clusters as descriptors of **author types**



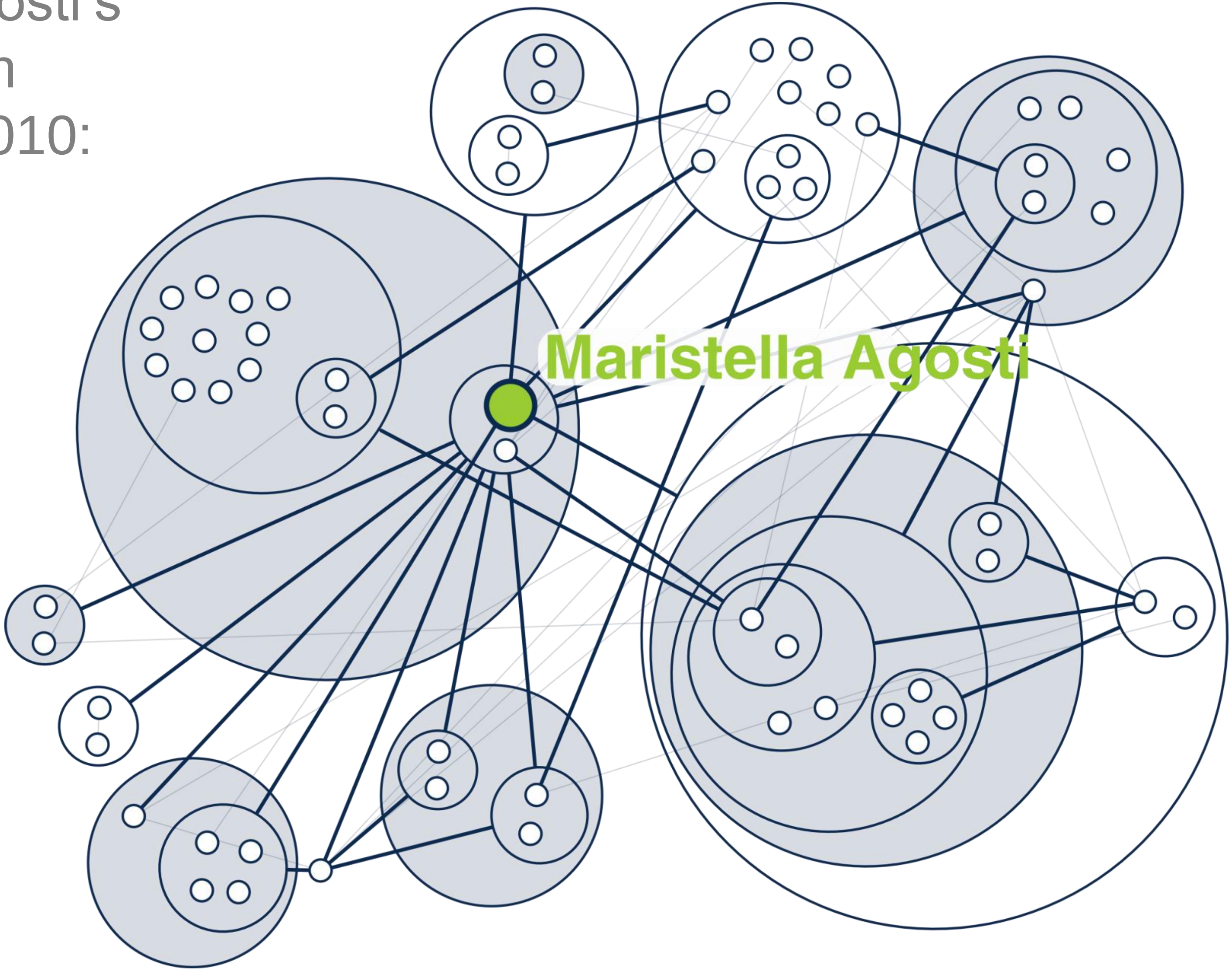
Methodology

1) Application of **Power Graph Analysis** to co-authorship graphs
2) Definition of **author features** based on power graph structure, number & impact of publications
3) Computation of **evolution indices (EI)**
4) **Clustering** based on EI
5) Employ EI of clusters as descriptors of **author types**

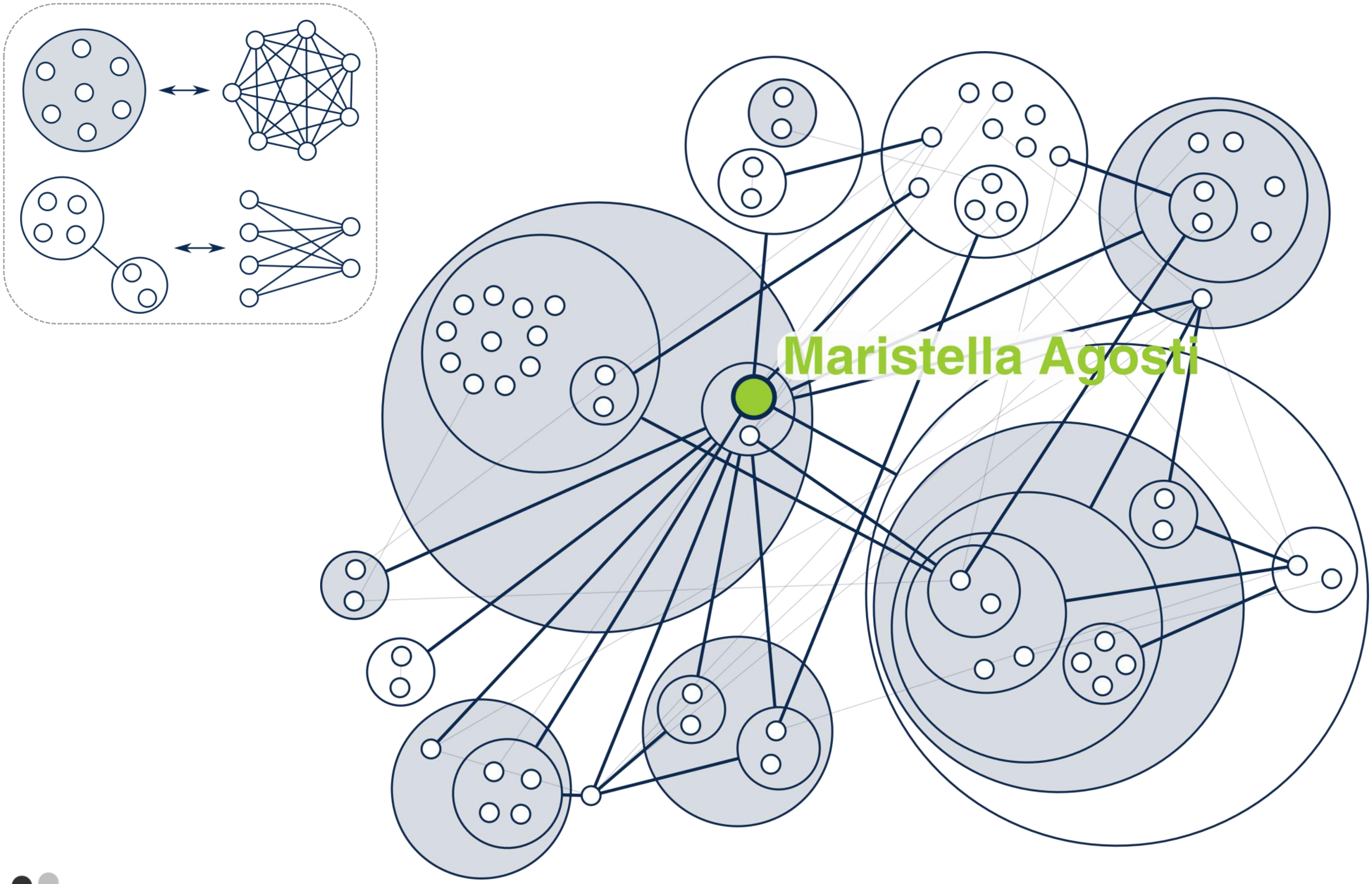


Power Graph Analysis

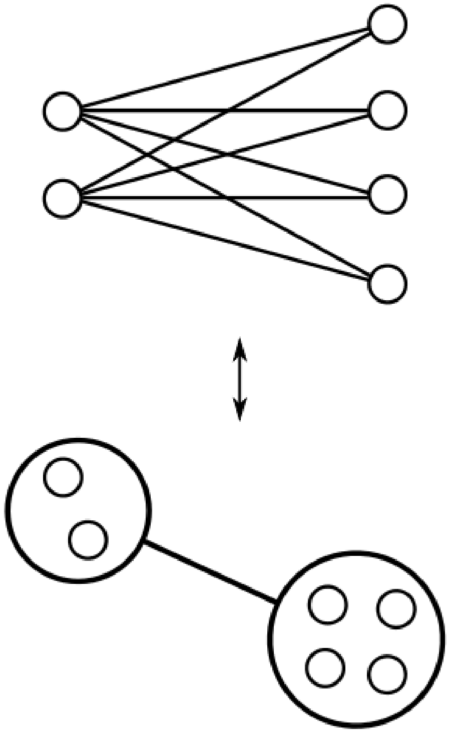
Maristella Agosti‘s Co-Authors in DBLP until 2010:



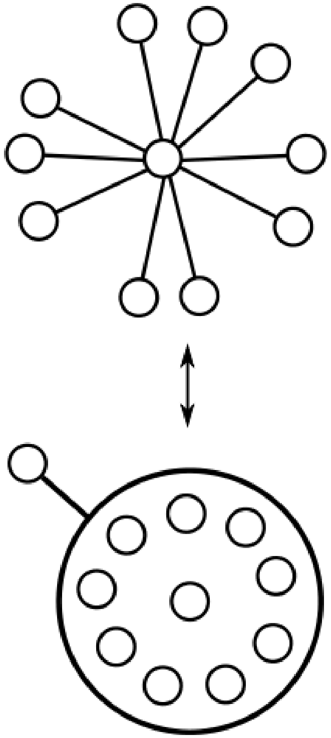
Power Graph Analysis



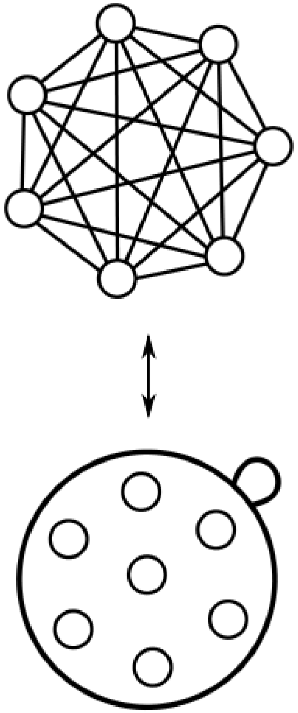
Power Graph Analysis



Biclique motif

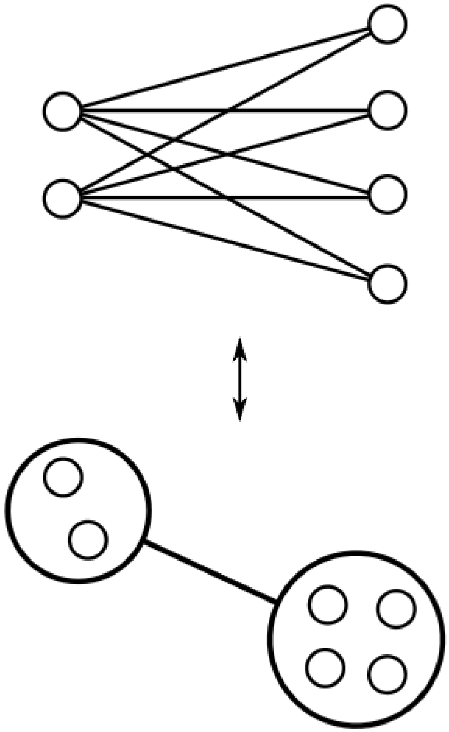


Star motif



Clique motif

Power Graph Analysis



Successfully applied in:

Biology

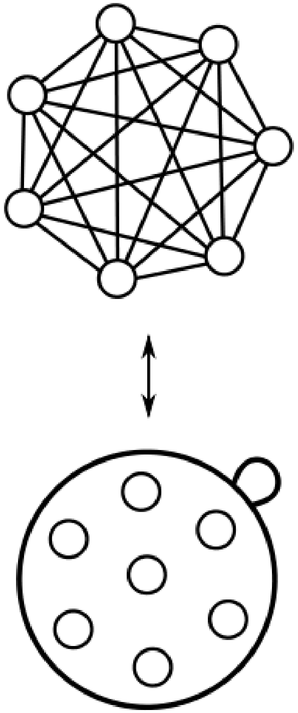
- Protein interaction networks
- Homogenity networks
- Genetic networks
- Regulatory networks
- Co-location networks

Social networks

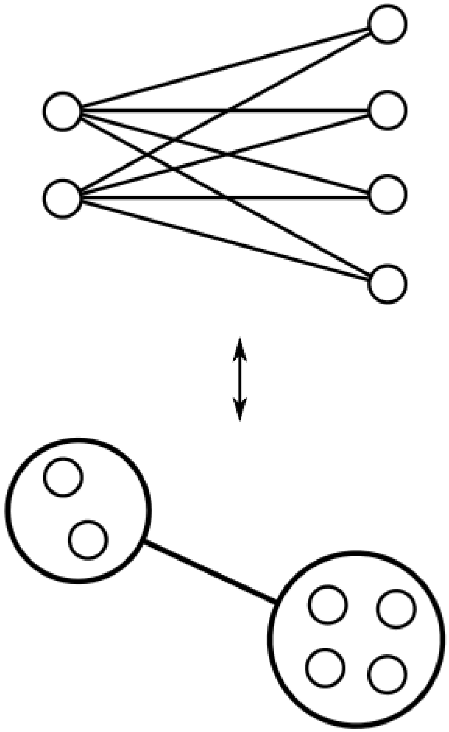
- Co-appearance networks

Computer science

- Class diagrams



Power Graph Analysis



Successfully applied in:

Biology

- Protein interaction networks
- Homogenity networks
- Genetic networks
- Regulatory networks
- Co-location networks

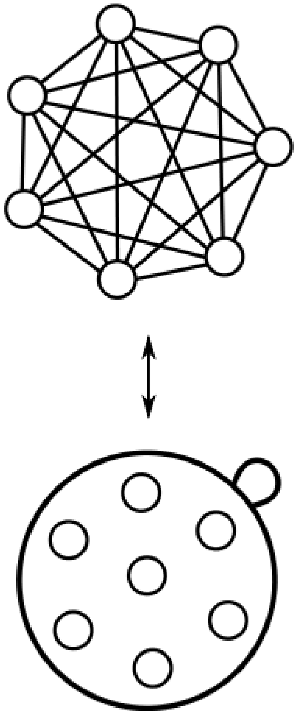
Social networks

- Co-appearance networks

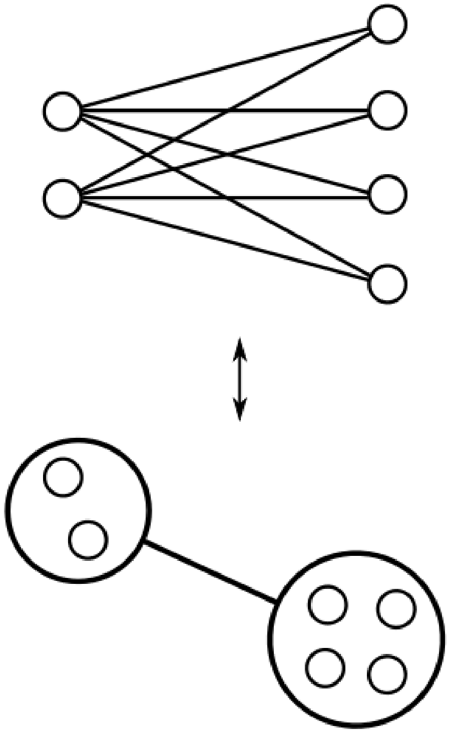
Computer science

- Class diagrams

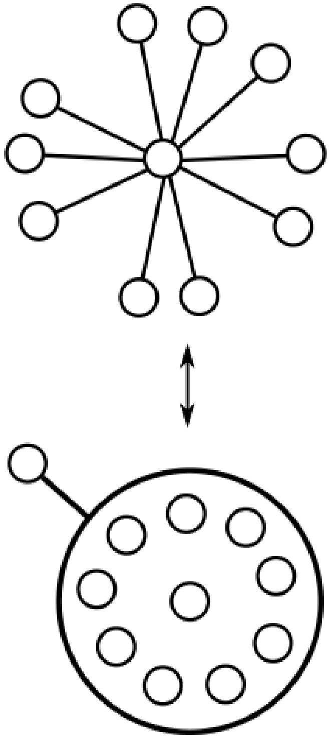
Rich in clique and/or bi-clique motifs



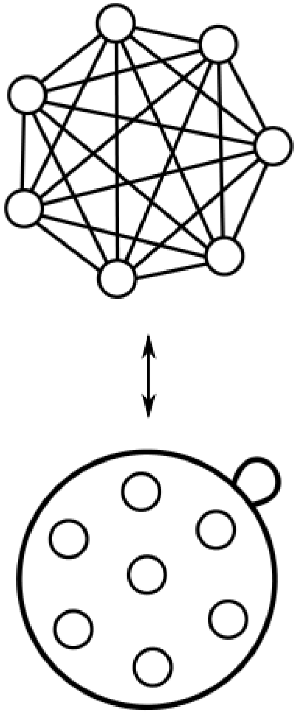
Power Graph Analysis



Biclique motif

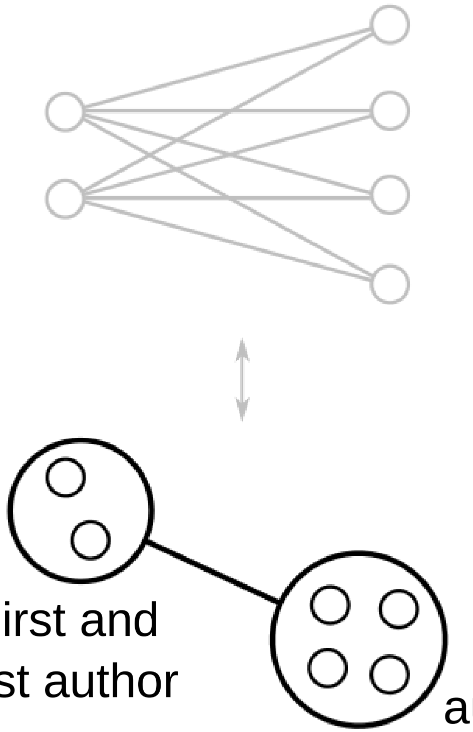


Star motif



Clique motif

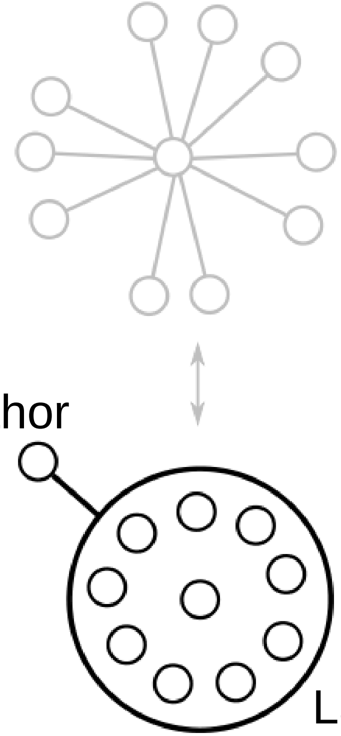
Co-author network creation



First and last author

Middle authors

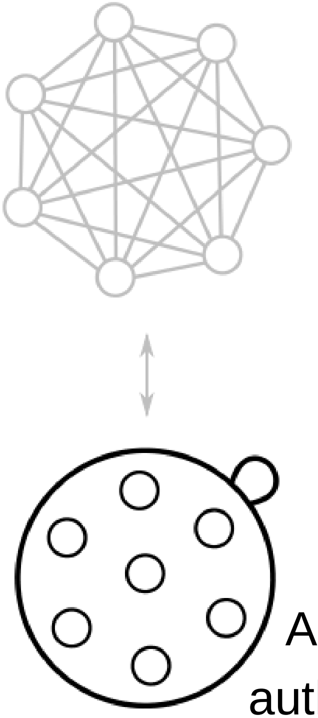
Biclique motif



First author

Second ... Last author

Star motif

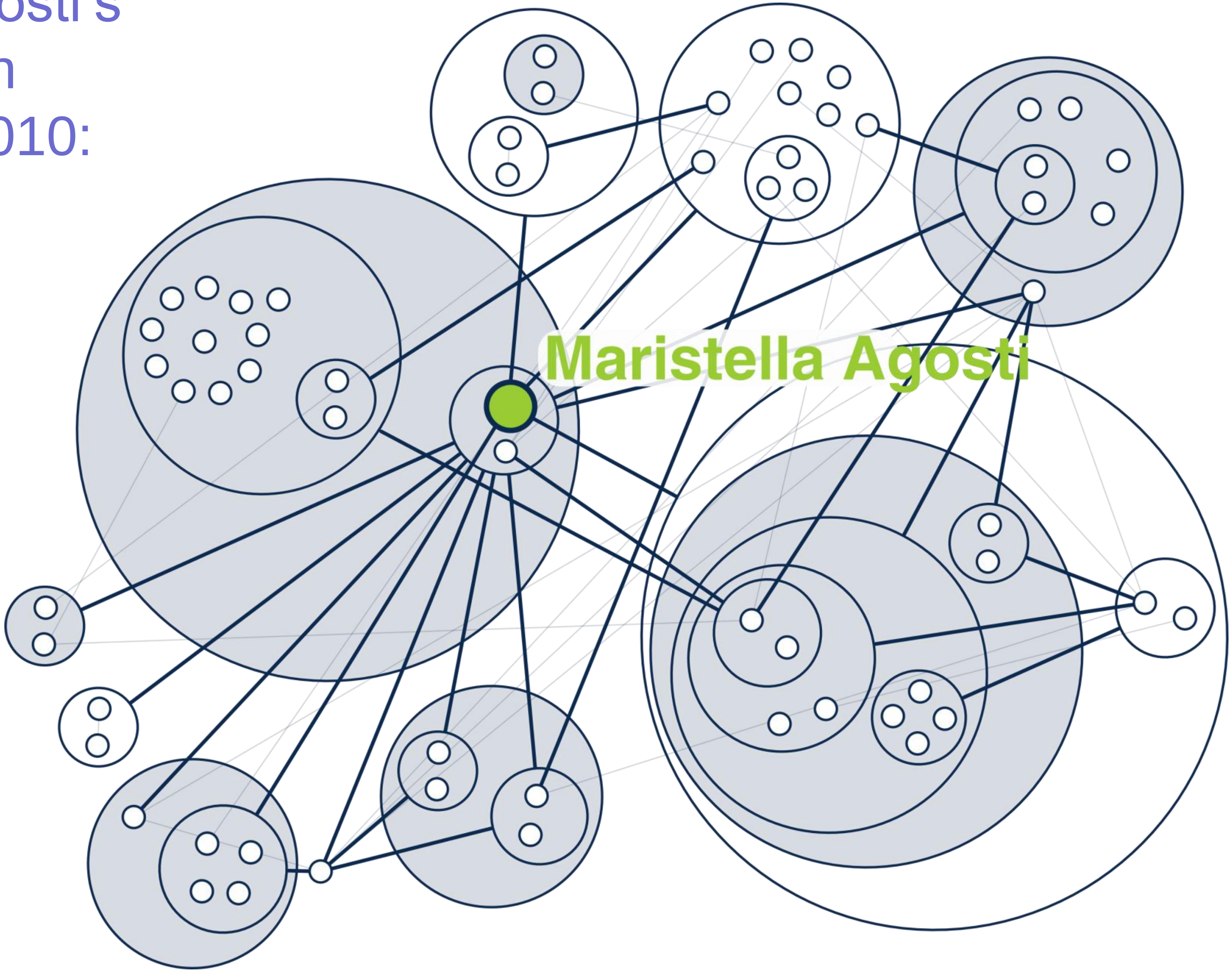


All co-authors

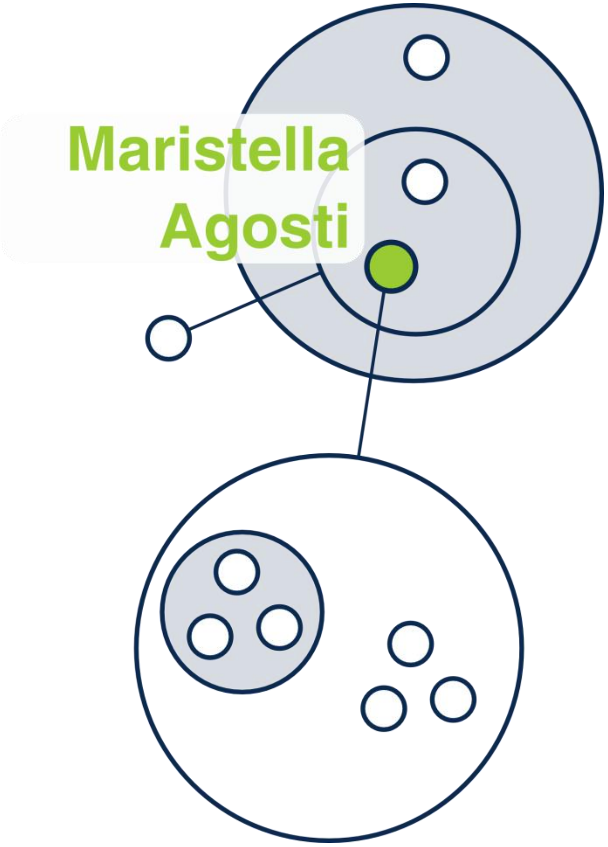
Clique motif

Power Graph Analysis

Maristella Agosti‘s Co-Authors in DBLP until 2010:

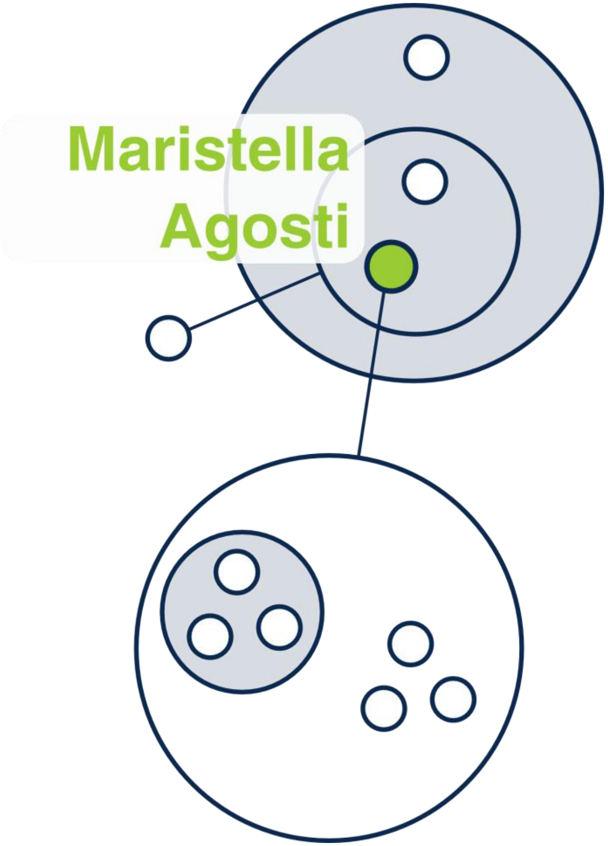


Co-authorship Evolution

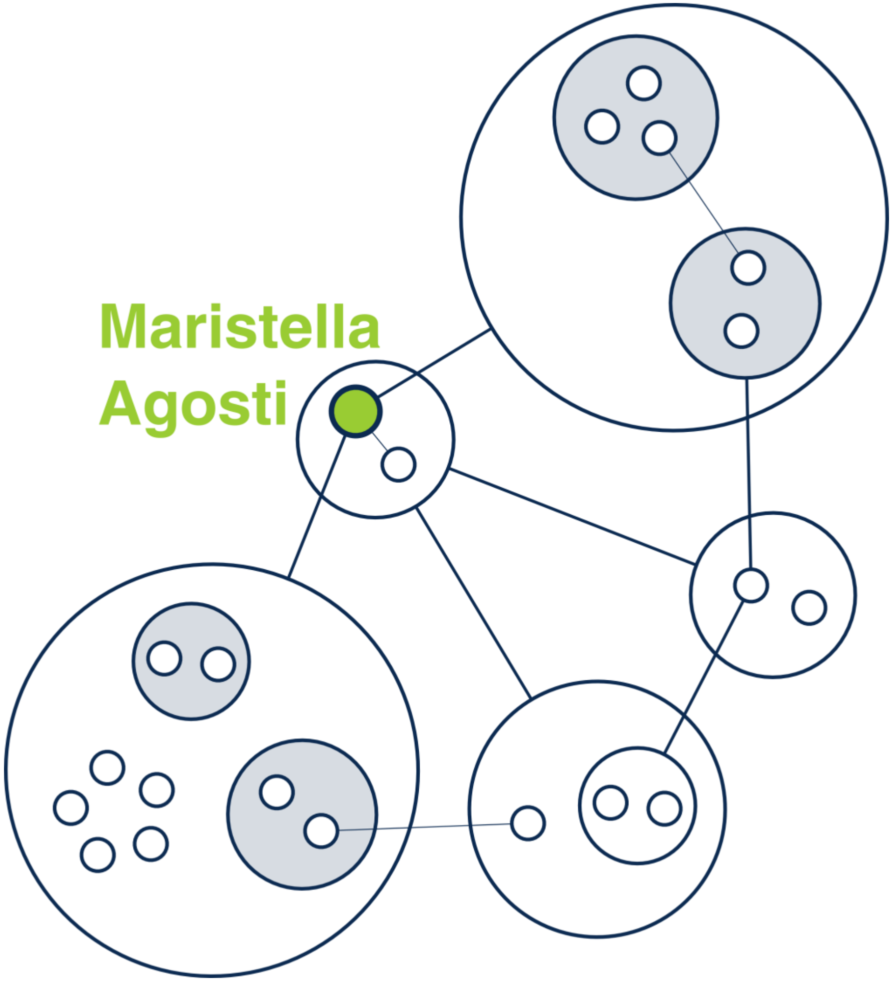


2000

Co-authorship Evolution



2000



2005

Methodology

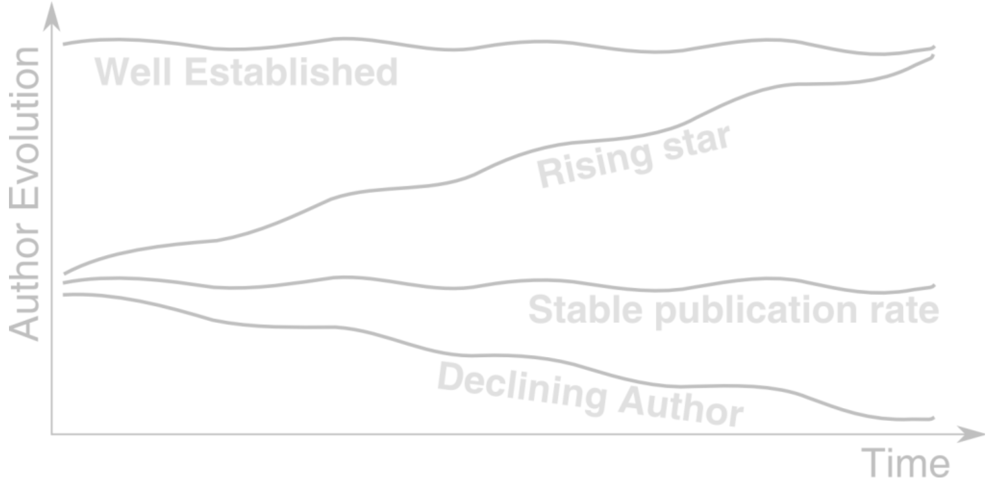
1) Application of **Power Graph Analysis** to co-authorship graphs

2) Definition of **author features** based on power graph structure, number & impact of publications

3) Computation of **evolution indices (EI)**

4) **Clustering** based on EI

5) Employ EI of clusters as descriptors of **author types**

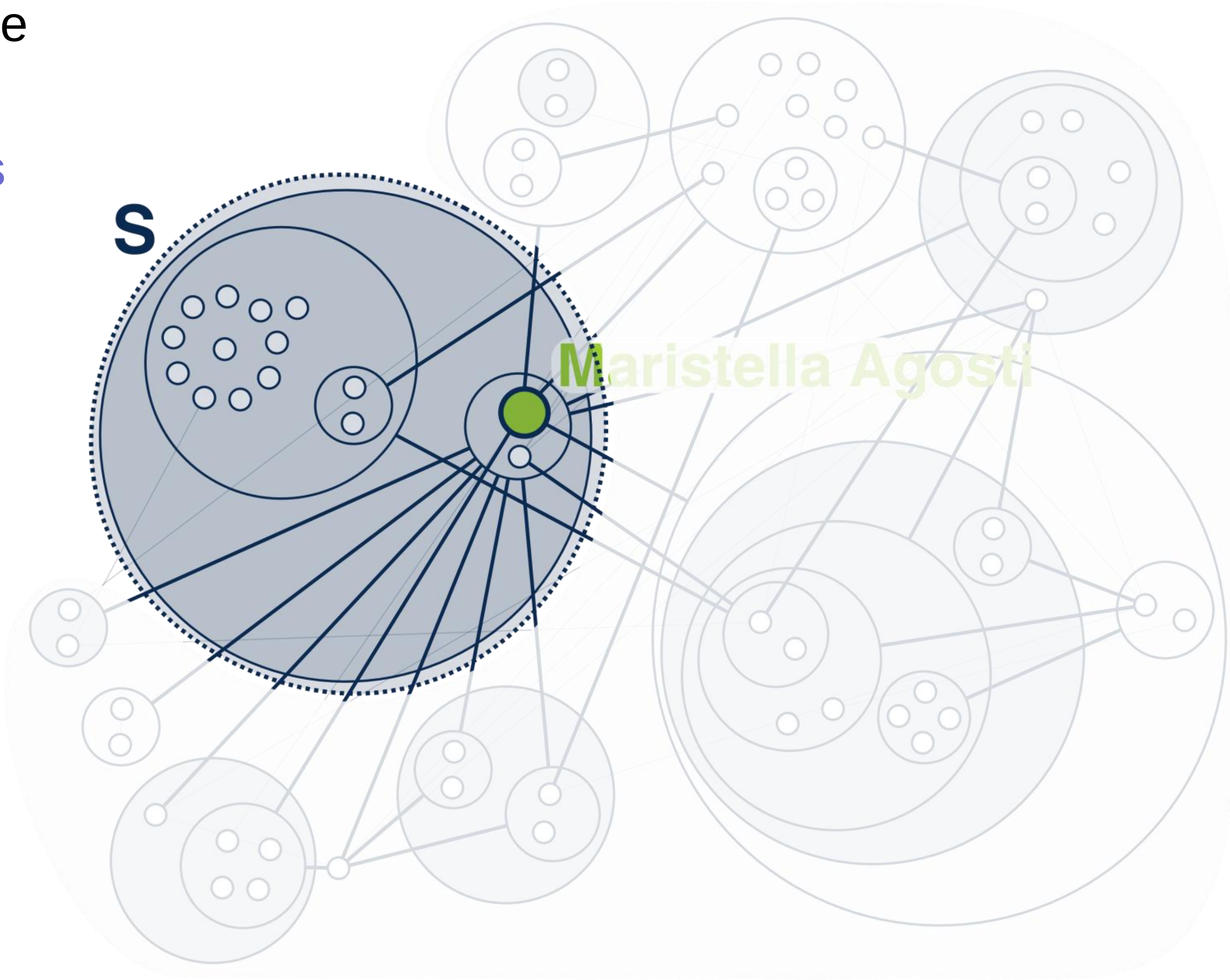


Author features

S – Size of Power Node

containing author:

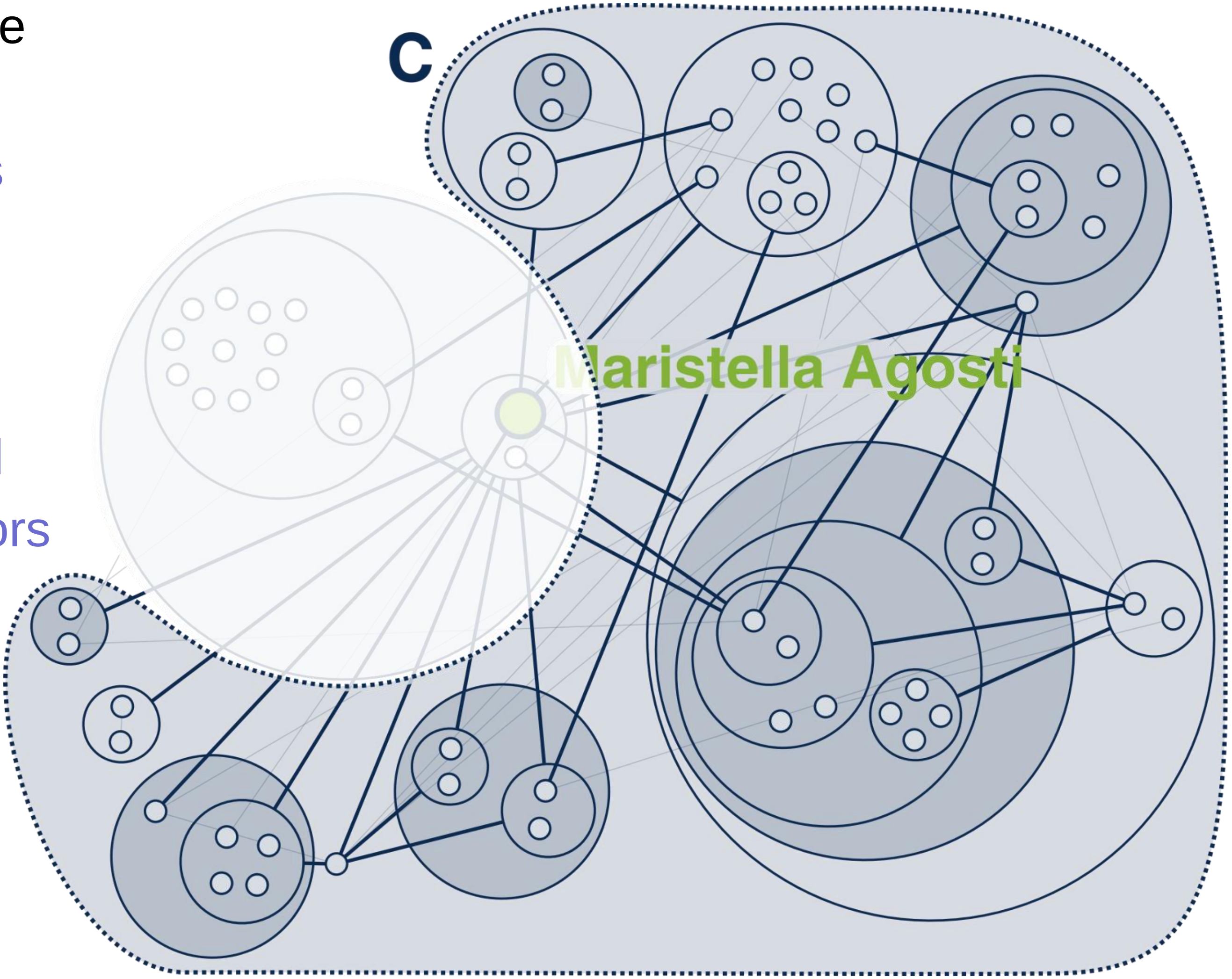
Regular co-authors



Author features

S – Size of Power Node
containing author:
Regular co-authors

C – Sum of sizes of
connected power
nodes: Occasional
& indirect co-authors



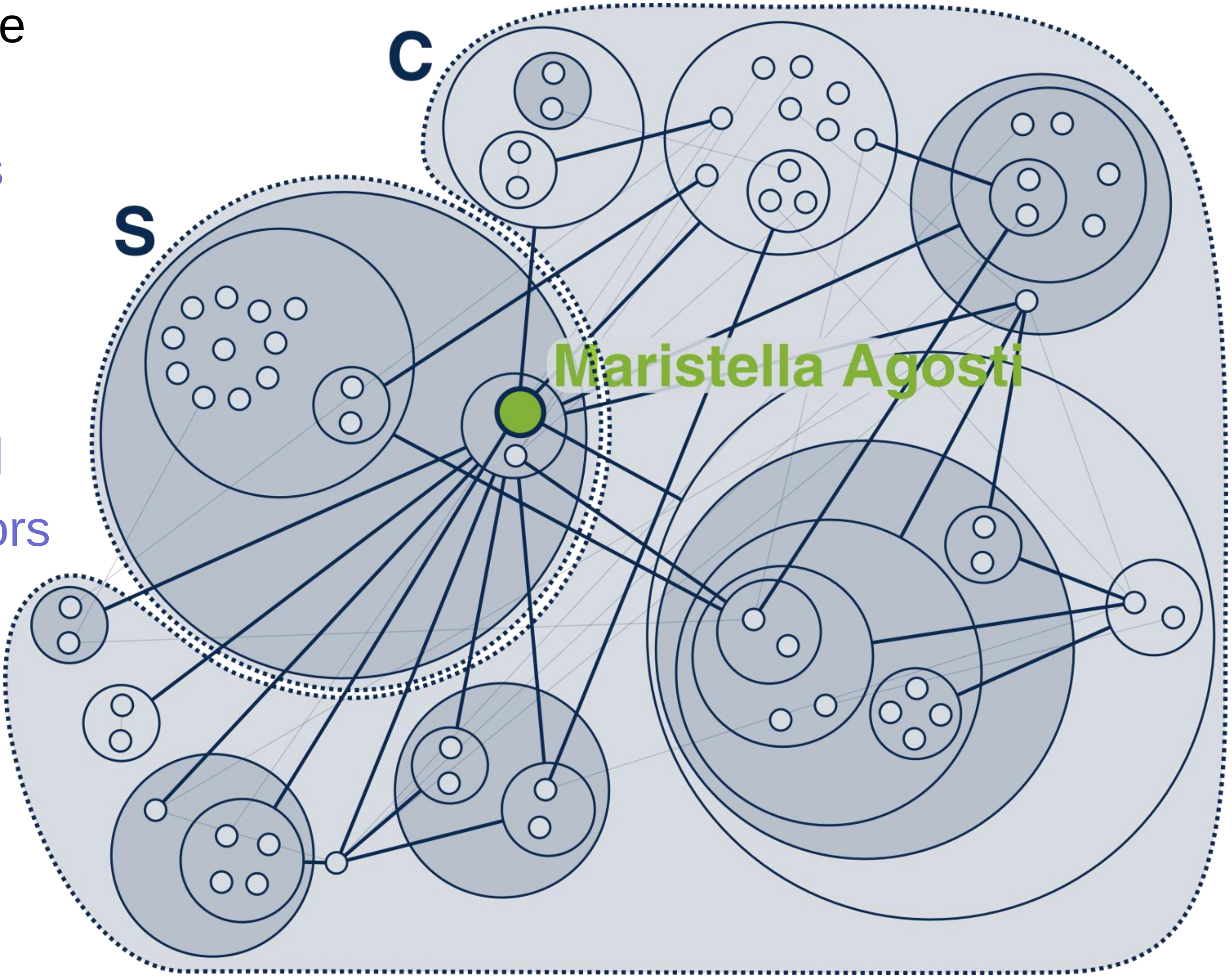
Author features

S – Size of Power Node containing author: Regular co-authors

C – Sum of sizes of connected power nodes: Occasional & indirect co-authors

P – Number of publications

I – Aggregated impact factor



Methodology

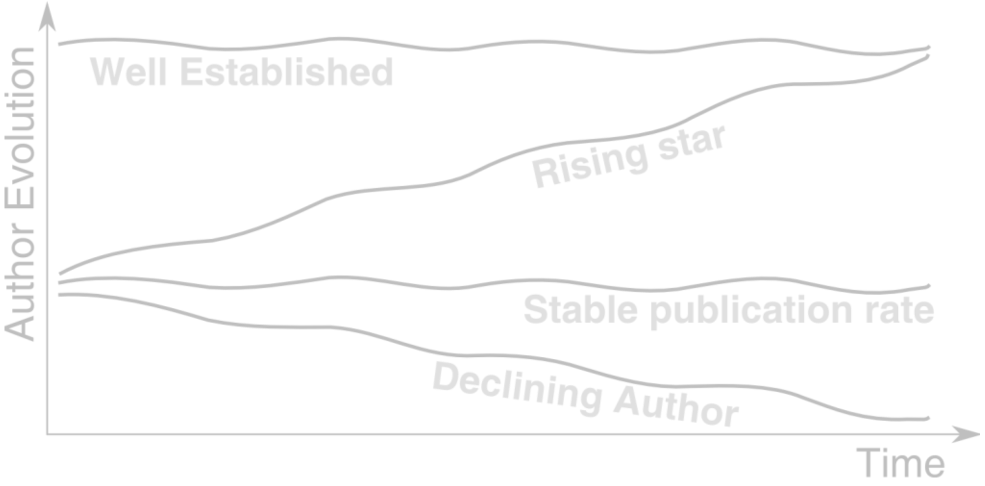
1) Application of **Power Graph Analysis** to co-authorship graphs

2) Definition of **author features** based on power graph structure, number & impact of publications

3) Computation of **evolution indices (EI)**

4) **Clustering** based on EI

5) Employ EI of clusters as descriptors of **author types**



Evolution Index

S – Size of Power Node containing author: Regular co-authors

C – Sum of sizes of connected power nodes: Occasional & indirect co-authors

P – Number of publications

I – Aggregated impact factor

For every time point change in feature:

$$C(S)_{ik} = \frac{S_{ik} - S_{(i-1)k}}{S_{ik}}$$

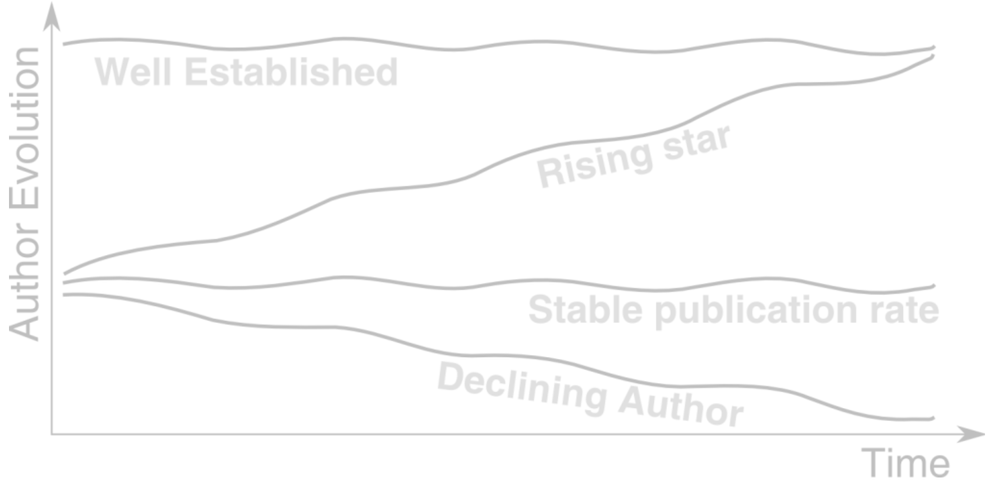
Evolution index:

$$EI(S)_{Tk} = \max_{t_i \in T} C(S)_{ik} \cdot S_{t_n k} \cdot \sum_{t_i \in T} C(S)_{ik}$$

Measures dynamics of author in S

Methodology

1) Application of **Power Graph Analysis** to co-authorship graphs
2) Definition of **author features** based on power graph structure, number & impact of publications
3) Computation of **evolution indices (EI)**
4) **Clustering** based on EI
5) Employ EI of clusters as descriptors of **author types**



Methodology

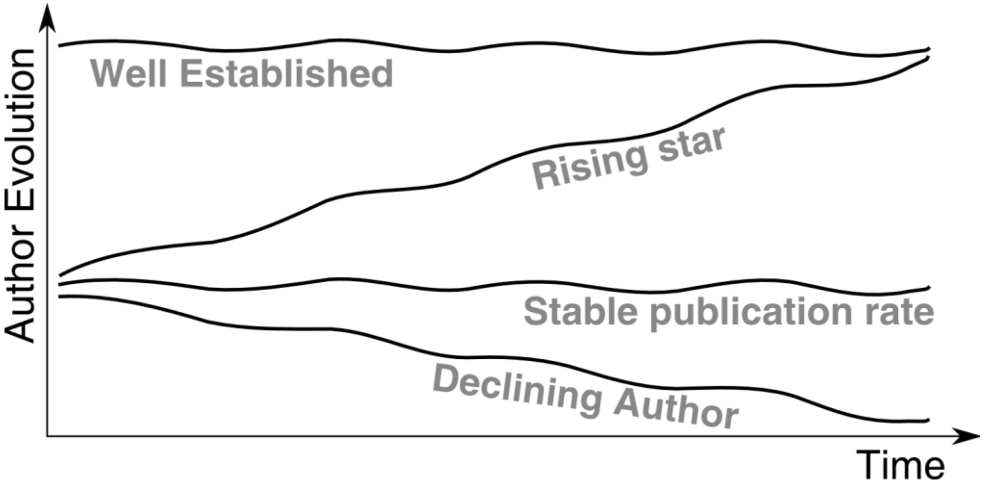
1) Application of **Power Graph Analysis** to co-authorship graphs

2) Definition of **author features** based on power graph structure, number & impact of publications

3) Computation of **evolution indices (EI)**

4) **Clustering** based on EI (bi-secting K-Means)

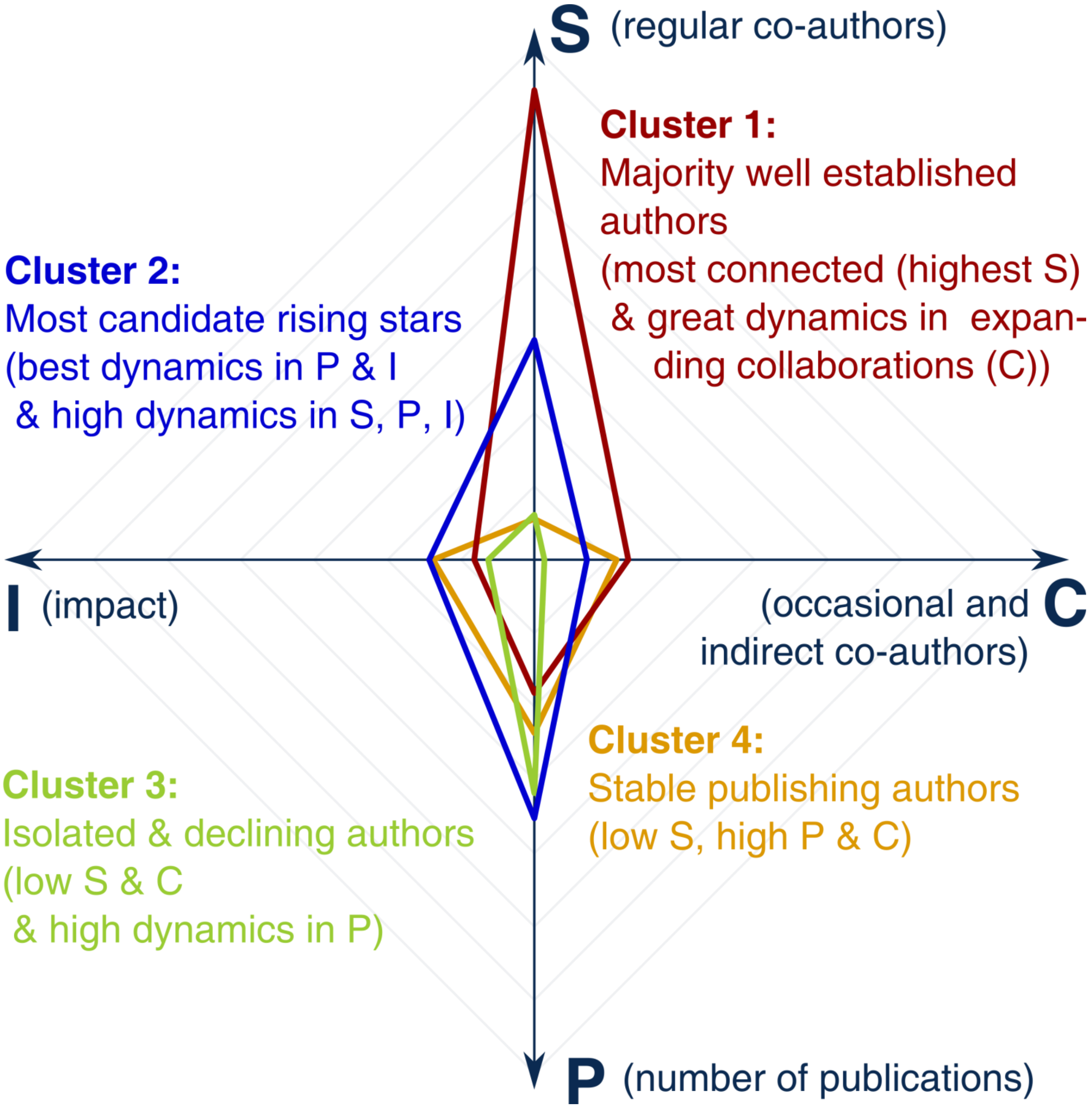
5) Employ EI of clusters as descriptors of **author types**



Results

On publication data from DBLP

(2000-2010)



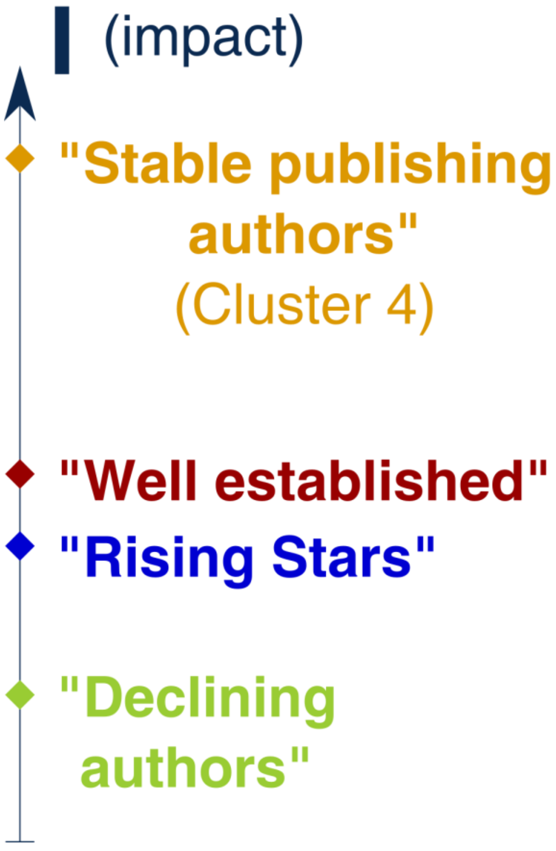
Cluster 1:
Majority well established authors
(most connected (highest S) & great dynamics in expanding collaborations (C))

Cluster 2:
Most candidate rising stars
(best dynamics in P & I
& high dynamics in S, P, I)

Cluster 3:
Isolated & declining authors
(low S & C
& high dynamics in P)

Cluster 4:
Stable publishing authors
(low S, high P & C)

Validation



I (impact)

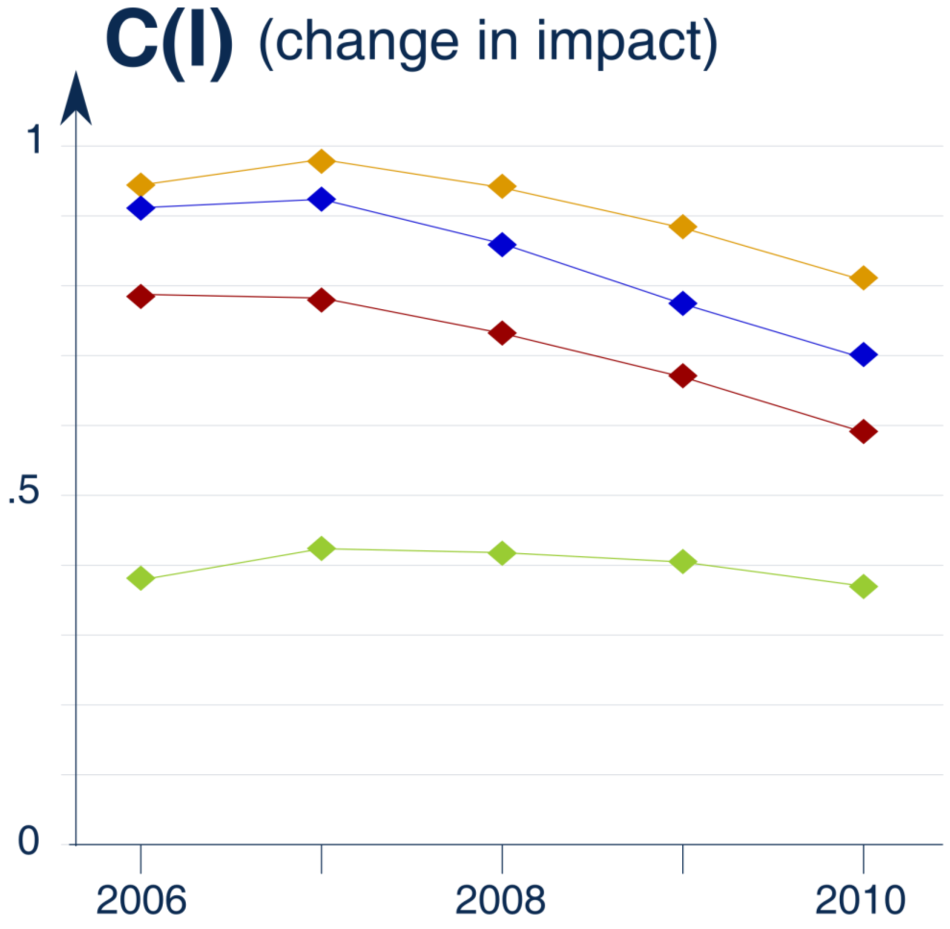
"Stable publishing authors" (Cluster 4)

"Well established"

"Rising Stars"

"Declining authors"

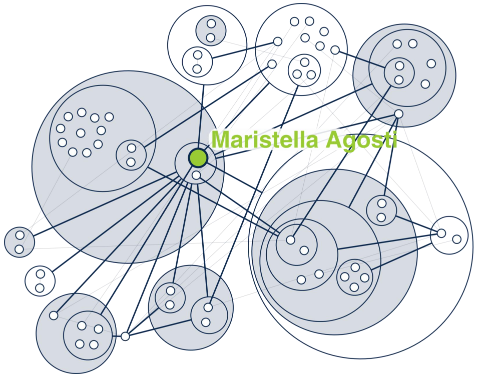
Impact of clustering for **2000-2005**

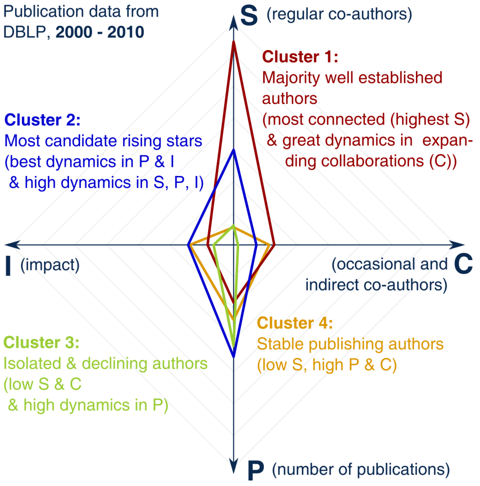


C(I) (change in impact)

Yearly increase in authors' impact factor for **2006-2010**

Conclusion



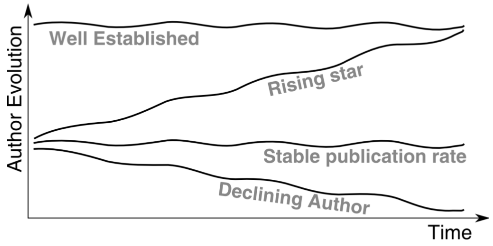

Publication data from DBLP, **2000 - 2010**

Cluster 1:
Majority well established authors
(most connected (highest S) & great dynamics in expanding collaborations (C))

Cluster 2:
Most candidate rising stars
(best dynamics in P & I & high dynamics in S, P, I)

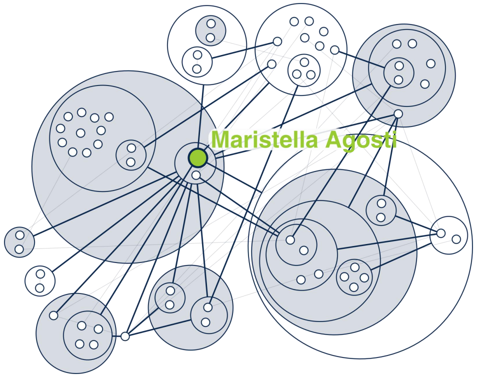
Cluster 3:
Isolated & declining authors
(low S & C & high dynamics in P)

Cluster 4:
Stable publishing authors
(low S, high P & C)

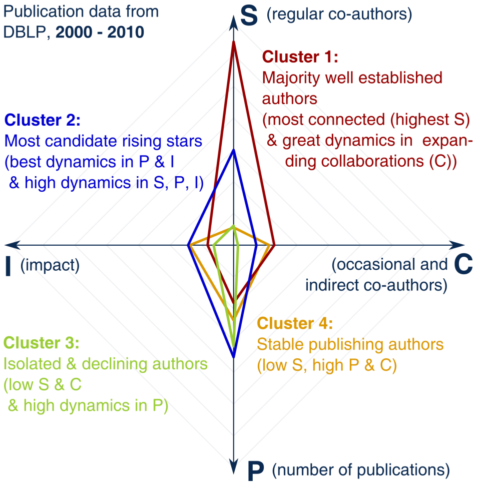


- **Novel methodology** for clustering authors based on graph mining
- **Fast and scalable** approach to monitor the evolution of authors over time
- Experimental validation in **real world data set** (DBLP – 120,000 distinct authors, **550,000 co-authorships**, 2,000 venues – 11 such graphs covering years 2000 - 2010)
- Analyzing authors as part of a wider science network supports modeling their profiles

Conclusion



Publication data from DBLP, **2000 - 2010**

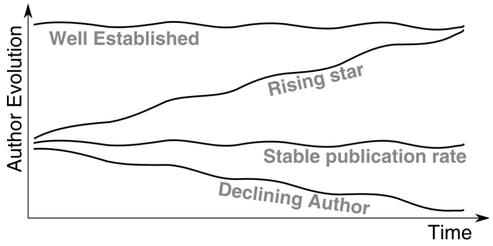


Cluster 1:
Majority well established authors
(most connected (highest S)
& great dynamics in expanding collaborations (C))

Cluster 2:
Most candidate rising stars
(best dynamics in P & I
& high dynamics in S, P, I)

Cluster 3:
Isolated & declining authors
(low S & C
& high dynamics in P)

Cluster 4:
Stable publishing authors
(low S, high P & C)



- Novel methodology for clustering authors based on graph mining
- Fast and scalable approach to monitor the evolution of authors over time
- Experimental validation in real world data set (DBLP – 120,000 distinct authors, 550,000 co-authorships, 2,000 venues – 11 such graphs covering years 2000 - 2010)
- Analyzing authors as part of a wider science network supports modeling their profiles

‘Science = Evolving network of people’

Acknowledgement

George Tsatsaronis

Iraklis Varlamis HAROKOPION UNIVERSITY

Sunna Torge

Kjetil Nørvåg NTNU – Trondheim Norwegian University of Science and Technology

Matthias Zschunke

George Tsatsaronis

Iraklis Varlamis

Power Graphs

Loic Royer

Michael Schroeder

biotec Biotechnology Center TU Dresden

TECHNISCHE UNIVERSITÄT DRESDEN

Loic Royer

Michael Schroeder