PRO-VE'10
11th IFIP Working Conference on VIRTUAL ENTERPRISES
Saint-Etienne, France, 11-13 October 2010

Eigenvector centrality based on shared research topics in a scientific community

Antonio P. Volpentesta, Alberto M. Felicetti

About us

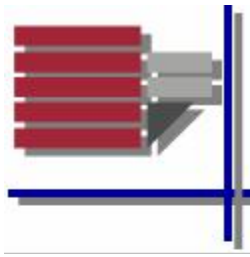
Prof. **ANTONIO VOLPENTESTA**

volpentesta@deis.unical.it

Dr. **ALBERTO FELICETTI**

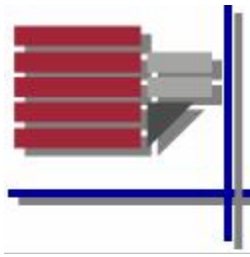
afelicetti@deis.unical.it

GIUDALab:
a lab of economics and management engineering
at **University of CALABRIA**



Outline

- *Introduction and Theoretical Background*
- *Eigenvector centrality for weighted multi-hypergraphs*
- *The weighted multi-hypergraph model*
- *A model instantiation for the Pro-VE community*
- *Conclusions and future works*



Outline

- *Introduction and Theoretical Background*
- *Eigenvector centrality for weighted multi-hypergraphs*
- *The weighted multi-hypergraph model*
- *A model instantiation for the Pro-VE community*
- *Conclusions and future works*

Introduction

Scientific Community (SC):

Networks of scientists, researchers and professionals who aim to produce, in a collaborative way, new knowledge within a specific domain or issue-area.

Key Factor for Scientific Knowledge development → ***COLLABORATION***

ISSUE:

Scholars are often not aware of other researchers who are working on similar projects and have same interests

Introduction

- Identify researchers who work on the same research topic.
- Highlight the ***"importance"*** of the different research topics (RTs) within a scientific community (SC).

What does it mean "importance"?

Relevance of a RT in a SC.

How do we can quantify "importance"?

By analyzing Centrality in a collaboration network of scientists.

What is centrality?

Network centrality: importance of a position within a network.

Idea of centrality comes from social networks literature.

Multiple meanings of centrality:

Degree centrality: Number of links incident upon a node.

Closeness centrality: Nodes are more central if they can reach other nodes 'easily.'

Betweeness centrality: Based on shortest paths in a network.

Eigenvector centrality: Important nodes have important "friends"

Introduction

Previous studies:

Based primarly on graph structures.

Few attempts in utilizing hypergraph in social network modeling.

What we propose:

Model based on weighted multi-hypergraph to represent relationship between researchers and research interests.

Measure of the importance in a such network, by extending eigenvector centrality notion to a weighted multi-hypergraph.

Outline

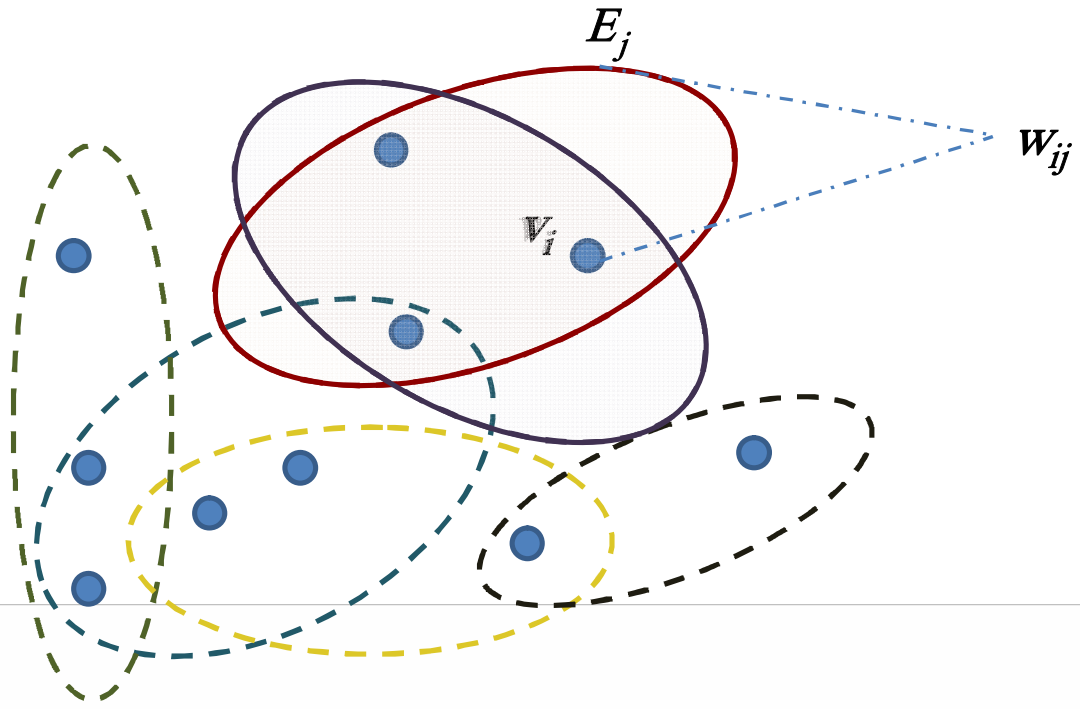
- *Introduction and Theoretical Background*
- *Eigenvector centrality for weighted multi-hypergraphs*
- *The weighted multi-hypergraph model*
- *A model instantiation for the Pro-VE community*
- *Conclusions and future works*

The weighted multi-hypergraph model

$\mathcal{H}(\mathcal{V}, \mathcal{E})$

$\mathcal{V} = \{v_1, \ldots, v_m\}$ is a set of vertices.

$\mathcal{E} = \{E_1, \ldots, E_n\}$ is a multi-set of nonempty subsets of $\mathcal{V}$, called hyperedges.



A.P. Volpentesta and A.M. Felicetti
GIUDALab – DEIS - University of Calabria, ITALY

Eigenvector centrality for weighted multi-hypergraphs

***Mutually reinforcing relationship* assumption:**

An important hyperedge is a hyperedge whose elements are important vertices;

An important vertex is a vertex that belongs to many important hyperedges.

Numerically:

$\mathbf{x_i}$ → 'importance' of vertex V_i

$$x_i = c_1 \sum_{j=1}^{n} w_{ij}\, y_j \,, \qquad \text{for i} = 1, \ldots, \text{m.}$$

$\mathbf{y_j}$ → 'importance' of hyperedge E_j

$$y_j = c_2 \sum_{i=1}^{m} w_{ij}\, x_i \,, \qquad \text{for j} = 1, \ldots, \text{n.}$$

In matrix notation with $x = (x_1, x_2, \ldots, x_m)$ and $y = (y_1, y_2, \ldots, y_n)$ this yields:

$$WW^t x = \lambda x, \qquad W^t W y = \lambda y, \qquad \text{where } \lambda = c_1 c_2.$$

A solution is given by setting $\lambda = \lambda^*$, the dominant W^tW's eigenvalue

A.P. Volpentesta and A.M. Felicetti
GIUDALab – DEIS - University of Calabria, ITALY

Outline

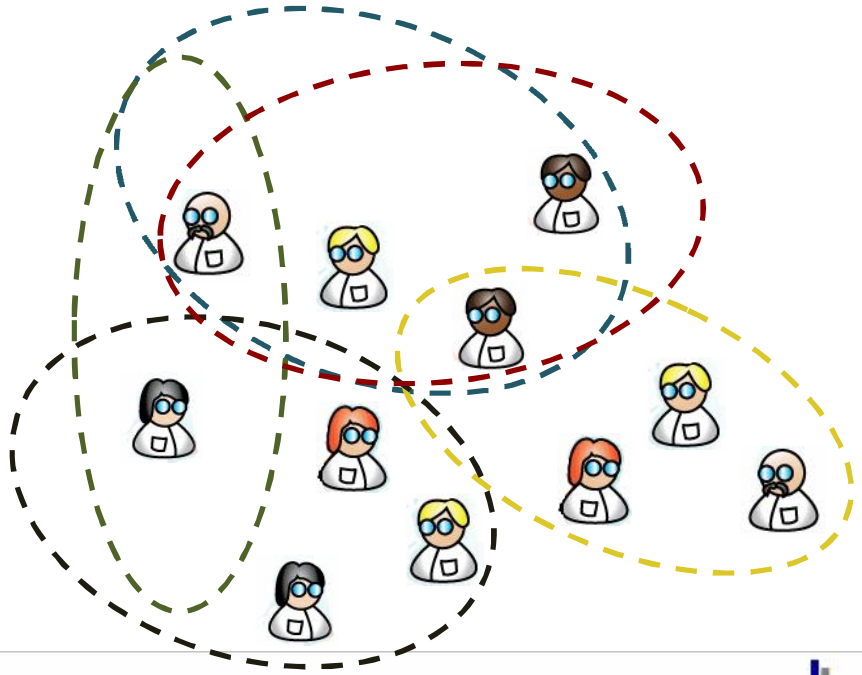
- *Introduction and Theoretical Background*
- *Eigenvector centrality for weighted multi-hypergraphs*
- *The weighted multi-hypergraph model*
- *A model instantiation for the Pro-VE community*
- *Conclusions and future works*

The weighted multi-hypergraph model

- $\mathbf{T} = \{t_1, \ldots, t_m\}$ an ordered set of research interests (research topics);
- $\mathbf{R} = \{r_1, \ldots, r_n\}$ an ordered set of researchers (authors), members of a SC;

Multi-hypergraph $\mathcal{H}\ (\mathcal{R}, \mathcal{E})$

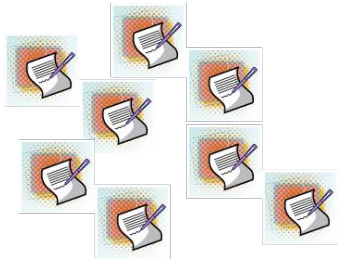
- $\mathcal{R} = \mathbf{R} = \{r_1, \ldots, r_n\}$;
- $\mathcal{E} = \{E_1, \ldots, E_n\}$, with $E_j = E(t_j)$



The weighted multi-hypergraph model

How to identify the research topics?

How to relate researcher to research topics?



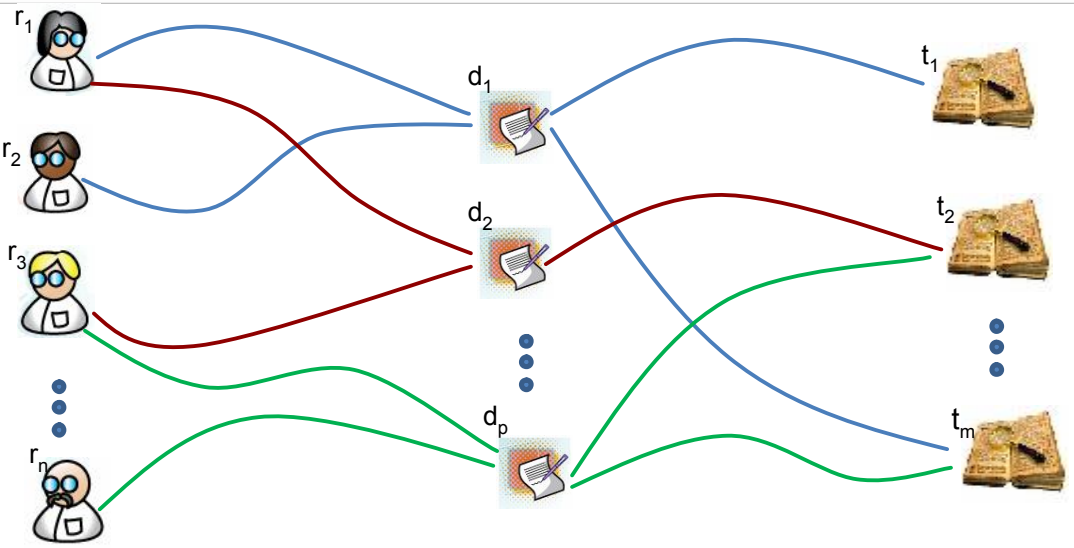
•D = $\{d_1,\ldots, d_p\}$ an ordered set of documents (scientific papers);

Assumptions:

Research interests of any researcher r_i are manifested on documents whose r_i is an author;

The relationship between researchers and interest groups may be derived through a semantic analysis of the documents' content.

The weighted multi-hypergraph model



- $\mathbf{A} \in \mathbb{R}^{n \times p}$ a binary matrix that represents the relationships between authors and documents produced by them.
- $\mathbf{B} \in \mathbb{R}^{p \times m}$ a nonnegative matrix that gives a measure of how much documents are devoted to research topics.
- $\mathbf{C} = (c_1, c_2, \ldots, c_p)$ a positive vector, where the generic entry c_k represents a measure of the popularity of d_k in the scientific community.

The weighted multi-hypergraph model

Assumptions:

- The content of a document is due in equal measure to all its authors: a_{ik}/h_k, (h_k , # of authors of d_k,) measures the document portion that is attributed to r_i.
- The number $b_{kj} \cdot c_k$ measures the contribution given by the research topic t_j to the popularity of the document d_k.
- The number $(a_{ik}/h_k)\cdot(b_{kj}\cdot c_k)$ measures the contribution given by the portion of d_k, dealing with t_j and attributed to r_i, to the popularity of d_k.

According to these assumptions and settings, we propose to estimate the weight associated to the couple (r_i, t_j), as follows:

$$w_{ij} = \sum_{k=1}^{p} (a_{ik}/h_k) \cdot (b_{kj} \cdot c_k)$$

The weighted multi-hypergraph model

- D = { d_1, d_2, d_3 }; T = { t_1, t_2, t_3, t_4 }; R = { r_1, r_2, r_3, r_4, r_5 }; C = (1, 2, 3);

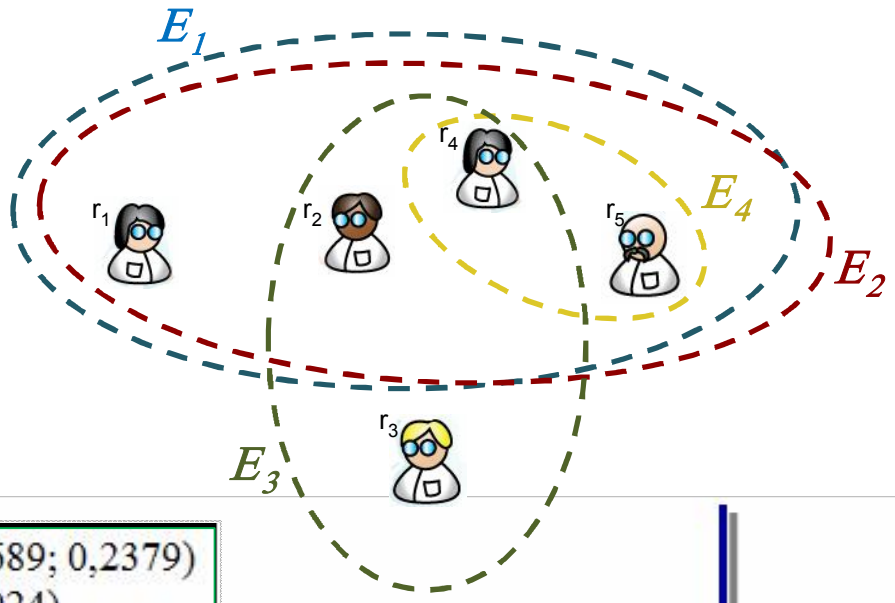
A	d_1	d_2	d_3
r_1	1	0	0
r_2	1	1	0
r_3	0	1	0
r_4	0	1	1
r_5	0	0	1

B	t_1	t_2	t_3	t_4
d_1	3/4	1/4	0	0
d_2	0	0	1	0
d_3	1/4	1/4	0	1/2

$$h_k = \frac{1}{\Sigma_i a_{ik}} \qquad k=1,2,3,$$

i.e. $(h_1, h_2, h_3) = (2, 3, 2)$;

W	t_1	t_2	t_3	t_4
r_1	3/8	1/8	0	0
r_2	3/8	1/8	2/3	0
r_3	0	0	2/3	0
r_4	3/8	3/8	2/3	3/4
r_5	3/8	3/8	0	3/4



Results:

x =(0,0656; 0,1966; 0,1309; 0,3689; 0,2379)
y =(0,2165; 0,1729; 0,3082; 0,3024)

Outline

- *Introduction and Theoretical Background*
- *Eigenvector centrality for weighted multi-hypergraphs*
- *The weighted multi-hypergraph model*
- *A model instantiation for the Pro-VE community*
- *Conclusions and future works*

A model instantiation for the Pro-VE community

- D: all selected papers that were published in the books of the last five Pro-Ve conferences (2005-2009);
- R: researchers who appeared as an author of at least one scientific article published in such books
- A represents relation between authorship and Pro-Ve papers.

What is a Research Topic (RT) in the Pro-Ve community?

We have modeled a research topic in the triple (OF, DA, ES) , where:

- OF is the set of Collaborative Network (CN) Organizational Forms.
- DA is the set of Dimensional Aspects (Derived from ARCON Model) of a CN.
- ES is the set of the economic sectors, each one encompassing real business environments, where CN principles are instantiated and implemented.

In other words, a RT is characterized by a dimensional aspect of a CN organizational form and possibly a case study or an application in primary industry, manufacturing, industrial services or intellectual services

A Research Topic in the Pro-VE community

Dimensional Aspects

Structural dimension	Actors / relationships
	Roles
Componential dimension	Hardware / software resources
	Human resources
	Information / knowledge resources
	Ontology resources
Functional dimension	Processes
	Auxiliary processes
	Methodologies
Behavioral dimension	Prescriptive behavior
	Obligatory behavior
	Constraints and conditions
	Contracts and cooperation agreements
Meta dimension	
External view	

CN Organisational Forms

CN	**Continuous production driven Network**	Supply chain
		Virtual Government
	Market Opportunity driven Network	Virtual Enterprise
		Virtual Organization
		Extended Enterprise
		Virtual Team
	Human breeding environments (Communities)	Virtual Team
		User's Community
		Community of Practices
	Organizational breeding environments	Industry Cluster
		Industrial District
		Business Ecosystem
		Collaborative Virtual Lab
		Disaster rescue Net

Economic Sector

Primary Economic Sector	(i.e.): Agriculture, Fishing, Forestry, etc…
Secondary Economic Sector	(i.e.): Automotive, Construction, Electronics, Mechanical, Textile, etc
Tertiary Economic Sector	(i.e.): Industrial Services, Commerce, Transportation, Hospitality, Maintenance, etc..
Quaternary Economic Sector	(i.e.): Banking, Consulting, Education, Government Services, Healthcare, etc…
No real world application	

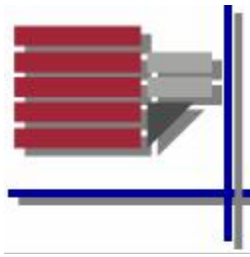
A model instantiation for the Pro-VE community

- Matrix **B:**

Obtained through a collaborative process of semantic analysis of Pro-VE papers;

- **Process** → Associate one or more instantiations of *(OF, DA, ES)* to any Pro-VE paper.

- **Assumptions**
  - Equi-distribution of the content of a paper among its research topics.
  - Any entry c_k is intantiated at x_k+1 (x_k: number of documents in D that cite d_k)



Outline

- *Introduction and Theoretical Background*
- *Eigenvector centrality for weighted multi-hypergraphs*
- *The weighted multi-hypergraph model*
- *A model instantiation for the Pro-VE community*
- *Conclusions and future works*

Conclusions

- We have proposed a model based on a weighted multi-hypergraph to represent relationships between researchers and research interests, grouping researchers with common interests.

- In order to measure the importance of researchers and research topics in a scientific community ,we have extended the eigenvector centrality notion to this general logical structure and we present an algorithmic approach.

- We have described a first application of the model to the Pro-VE Community.

Further developments

- Complete the implementation phase (we are collecting and validating data derived from a semantic analysis of Pro-Ve papers).
- Provide measurements and statistical analysis of the centrality of researcher and research topics within the Pro-Ve community.
- Extend this studies in order to map competences in a SC.
- Provide a competence map of the Pro-VE community.

Thanks for your attention

alberto.felicetti@deis.unical.it