Social Networks and Social Information Sharing Using Fuzzy Methods

Ronald R. Yager

Machine Intelligence Institute
Iona College
New Rochelle, NY 10801
yager@panix.com

Rapidly Emerging New Environment

- Interactive Web 2.0 applications

  LinkedIn / Facebook / Twitter / MySpace

- Cyberwarfare
- Exponential Increase in Universal Interconnectivity

We need technologies to help us better navigate this new environment.

Intelligent Social Network Modeling

Intelligent Social Network

Modeling and Computing Requires

Communications, Cooperation

and Coordination Between Man

and Machine

Major Difficulty

Human Beings Communicate, Understand and Reason Most Comfortably Using Linguistic Concepts

Machines Communicate, Understand and Reason Using Formal Mathematical Structures

The Success of Intelligent Social Network Modeling Requires us to Bridge this Gap

Bridging the Gap

- Should be Human Focused
- Communal Vocabulary
- Enable Machines to Comprehend and Manipulate Linguistic Concepts
- Linguistic Concepts are Granules

Communal Vocabulary

- Collection of Terms Commonly Understood by both Man and Machine
- Inter-Species Communication Uses Vocabulary
- Man Uses Linguistic Term
- Machine Uses Fuzzy Set Representation
- Man Determines the Content of the Vocabulary

Example Vocabularies

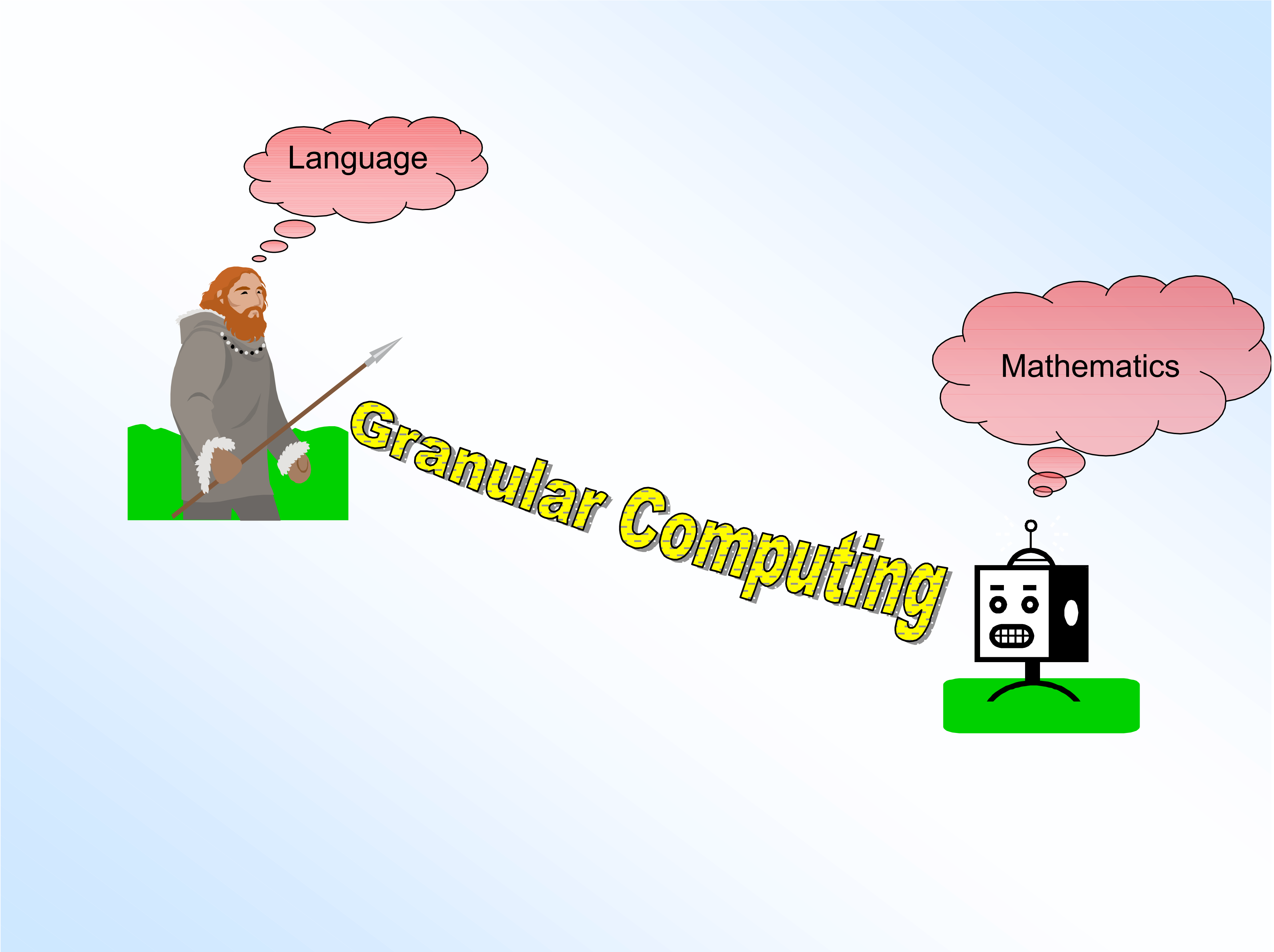
- Age
  {young, old, senior, kid, 23, “about 40”}
- Weather
  {cold, warm, swimming, 30°, nice}
- Proportions

  {most, some, “about half”, large, 35%}

Granular Computing is a Collection of <u>Set Based</u> Technologies that Allow for the Formal Representation and Manipulation of Human Focused Linguistic Concepts

Granular Computing Technologies

- Fuzzy Set Theory
- Dempster-Shafer Belief Structures
- Rough Sets
- Probabilistic Reasoning
- Possibility Probability Granules



Language
Granular Computing
Mathematics

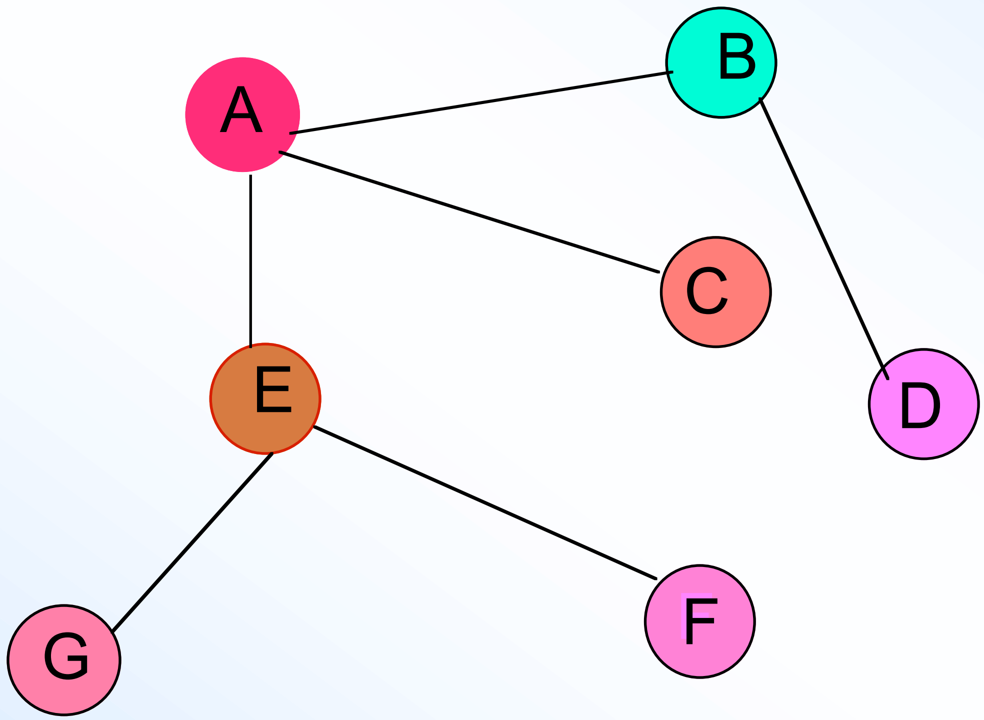
Role of Granular Computing in Social Networks

Extend the Capabilities for Analyzing Social Relational Networks by Enabling the Use of Human Like Concepts With Fuzzy Set and Granular Technologies

Inspiration and Motivation

- Formal Representation of Social Network is a Set Theoretical Object
- Granular Computing is Set Based Technology
- A Marriage between the Two is Natural

Relational Social Network



Communal Vocabularies for Social Network Analysis

- **Path Length**
  {long, short, moderate, 10 links}
- **Strength of Connection**
  {strong, weak, full}
- **Centrality**
  {center, margin, periphery}

Description of Network

- **Set of Nodes**

  $X = \{A, B, C, D, E, F, G\}$

- **Collection of Edges**

  $$E = \{(A, B), (A, C), (A, E), (B, D), (E, F), (E, G)\}$$

Mathematical Model of Network

- Set of Nodes X
- Relationship R on $X \times X$

  $R(x, y) = 1$ if Link from x to y

  $R(x, y) = 0$ if No Link from x to y
- R is a Subset of $X \times X$

Path in Social Networks

- Sequence of Nodes

  $x_1\ x_2\ x_3\ \ldots\ldots x_n$

- Sequence is a Path from x_1 to x_n if

  $\mathrm{Min}_{i=1 \text{ to } n-1}[R(x_i, x_{i+1})] = 1$

- Length of path(# of links) = n - 1
- Geo(x,y) = Length of Shortest x-y path

Composition of Relations

- R is a relation on $X \times X$
- $R(x, y) \in [0, 1]$
- Composition : $R^2 = R \blacklozenge R$
- $R^2(x,z) = Max_y(Min(R(x,y), R(y,z))$
- R^k is Composition K times: $R \blacklozenge R \blacklozenge R \blacklozenge R \blacklozenge R$
- R^k is a subset of $X \times X$

Paths and Composition

- $R^k(x, y) = 1$ if there exists a path of at most k links between nodes x and y

- Geo(x,y) is the smallest k such that $R^k(x, y) = 1$

Example of Using

Granular Computing for

Human Focused Analysis of

Social Network

Cliques and Clusters

- Subset S of X is called a Clique of order k if for all x, y $\in$ S we have Geo(x, y) $\leq$ k

- For all z $\notin$ S we have Geo(x, z) > k for some x $\in$ S

Human Definition of Clique

A subset S of nodes in the network is a clique if most of the elements in S are closely connected, none of the nodes in S are to far from each other and no element not in the clique is better connected to the members of the clique then any element in the clique.

Criteria for Clique

- C1: Most of the elements in S are closely connected
- C2: None of the elements in S are too far from the other elements
- C3: No element not in S is better connected to the members of the clique then any member of the clique

Determination of Cliqueness of S

- Obtain Degree of satisfaction of C1 by S
- Obtain degree of satisfaction of C2 by S
- Obtain degree of satisfaction of C3 by S
- Cliqueness of S is fusion of these values

$$Clique(S) = Min[C1(S), C2(S), C3(S)]$$

Satisfaction of C1

Most of elements in S are **Closely** connected

- Extract from communal vocabulary meaning of **Close** and **Most**
- **Close**: Fuzzy set Q where Q(k) is degree k links considered close
- **Most**: Fuzzy set M where M(p) is degree proportion p satisfies most

Closeness of Two Nodes

- Assume x and y in S
- $Close(x, y) = Max_k[Q(k) \wedge R^k(x, y)]$
- Marriage of Network Model & GC
  - Linguistic term Q
  - Set Representation of Network R^k

Calculation of $C_1(S)$

- Assume n_S is number of elements in S
- For each x_i in S calculate

$$p_i = \frac{1}{n_s - 1} \sum_{j \neq i} \mathrm{Close}(x_i, x_j)$$

- Using this we obtain

$$C_1(S) = \operatorname*{Max}_{x_i \in S}[M(p_i)]$$

Congestion in Networks

Uncongested Paths

- Number of arcs incident upon a node can interfere with performance of the node in a path
- Formulate idea of an uncongested path
- Useful in social networking systems such as **LinkedIn**
- Contact a person using connections via other people

Congested Node

- Node is a congested node if it has *many* incidence nodes
- Express as a fuzzy set Cong over the set X of nodes
- $\mathrm{Cong}(x_i) \in [0, 1]$ is degree to which x_i is a congested node
- Using Fuzzy Subset MANY

$$\mathrm{Cong}(x_i) = \mathrm{MANY}\Big(\sum_{j \neq i} R(x_i, x_j)\Big)$$

More General Formulation

- Define congested node using a Takagi-Sugeno type fuzzy systems model.
- V the number of incident arcs on a node.
- Cong indicate the degree to which node is congested node
- Use rules

If V is A_j then Cong is α_j

.

$$\mathrm{Cong}(x_i) = \frac{\sum_j A_j(q_i)\,\alpha_j}{\sum_j A_j(q_i)} \qquad \text{where } q_i = \sum_{k \neq i} R(x_i, x_k)$$

Basic Definition of Uncongested Path

- A path in which **all** intermediary nodes are uncongested.

- Sequence $\rho = x_1\, x_2\, ...\, x_q$.

- Degree ρ provides an uncongested path from x_1 to x_q.

$$\mathrm{Uncong}(\rho) = \underset{i=1 \text{ to } q-1}{\mathrm{Min}} [R(x_i, x_{i+1})] \wedge \underset{i=2 \text{ to } q-1}{\mathrm{Min}} [(1 - \mathrm{Cong}(x_i))]$$

More Sophisticated Definitions

Using Computing with words we can implement other formulations

Most of the nodes are uncongested

None of the nodes are very congested

Duration

- Generalization idea of length of the path
- Considers impact of congestion on the length path
- Sequence of nodes $\rho = x_1\, x_2 \ldots x_q$.

$$\mathrm{Dura}(\rho) = \sum_{j=1}^{q-1} \frac{1}{R(x_j, x_{j+1})} + \sum_{j=2}^{q-1} \left(\frac{1}{(1 - \mathrm{Cong}(x_j))} - 1 \right)$$

Special Cases

- Standard case do not consider congestion : Cong $(x_i) = 0$.

$$\mathrm{Dura}(\rho) = \sum_{j=1}^{q-1} \frac{1}{R(x_j, x_{j+1})}$$

- If ρ is a true path then $R(x_j, x_{j+1}) = 1$ for all j

  $\mathrm{Dura}(\rho) = q - 1$, the # of links in path.

- If any of $R(x_j, x_{j+1}) = 0$, the sequence ρ is not a path

  $\mathrm{Dura}(\rho) = \infty$

Situation with Some Congestion.

$$\frac{1}{1-\mathrm{Cong}(x_i)}-1=\frac{\mathrm{Cong}(x_i)}{1-\mathrm{Cong}(x_i)}$$

$$\mathrm{Dura}(\rho)=\#\text{ of links}+\sum_{j=2}^{q-1}\frac{\mathrm{Cong}(x_j)}{1-\mathrm{Cong}(x_j)}$$

$$\frac{\mathrm{Cong}(x_j)}{1-\mathrm{Cong}(x_j)}$$ is node **Delay**

Weighted Social Relational Networks

- $R(x, y)$ Strength of Connection
- $R(x, y) \in [0, 1]$
- Fuzzy Relation
- Composition : $R^2 = R \blacklozenge R$
- $R^2(x,z) = \mathrm{Max}_y(\mathrm{Min}(R(x,y), R(y,z))$
- R^k is Composition K times: $R \blacklozenge R \blacklozenge R \blacklozenge R \blacklozenge R$

Paths in Weighted Social Networks

- Sequence of Nodes

  $x_1\, x_2\, x_3 \ldots\ldots x_n$

- Strength of Path from x_1 to x_n

  $\mathrm{Min}_{i=1 \text{ to } n-1}[R(x_i, x_{i+1})]$

- Length of Path

$$\sum_{i=1}^{n-1} \frac{1}{R(x_i, x_{i+1})}$$

Linguistically Weighted Social Networks

- $R(x,y) \in$ **L**

- **L** = {no, very weak, weak, moderate,strong, very strong, full}

- Precisiation of Elements in L

  Ordinal Scale

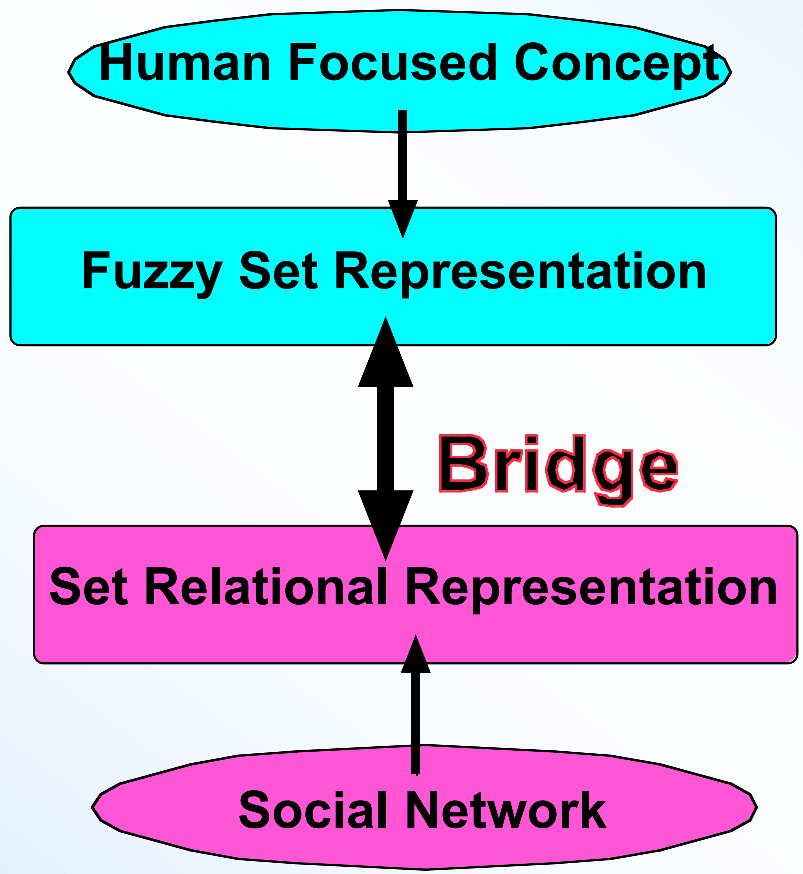
  Define elements as fuzzy sets of [0,1]

Intuitionistically Weighted Social Networks

- $R(x, y) = (\mathbf{a}, \mathbf{b})$
- $\mathbf{a}$ & $\mathbf{b} \in [0, 1]$
- $\mathbf{a} + \mathbf{b} \leq 1$
- **a** is degree of support for connection
- **b** is degree of support for no connection

Fuzzy Sets and Related Granular Computing Technologies Provide Fundamental Technologies for Social Network Analysis

Paradigm for Intelligent Social Network Analysis PISNA



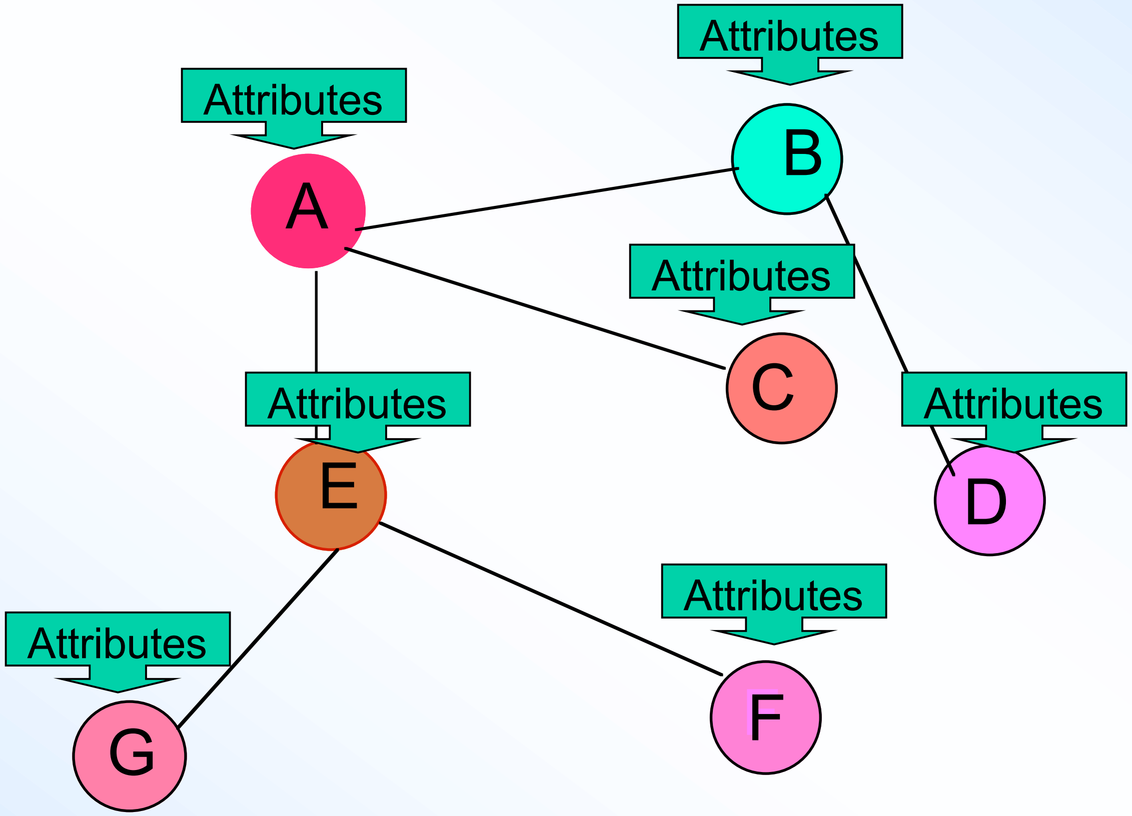
Vector Valued Nodes and Social Network Databases

Relational Network & Databases

- Weighted network <X, R>
- Each node has associated vector of attribute values
- View as combination of Relational network and database
- Web social networks: LinkedIn & Facebook
- Terror networks and criminal networks..

Structure of Database

- A collection of q attributes $U_1, ..., U_q$
- $U_i(x_j)$ denotes attribute U_i in case of node x_j
- Each $U_i(x_j)$ takes value from a domain Y_i.
- Each node has an associated q vector
- i^{th} component is value of i^{th} attribute for that node



Network View

Database View

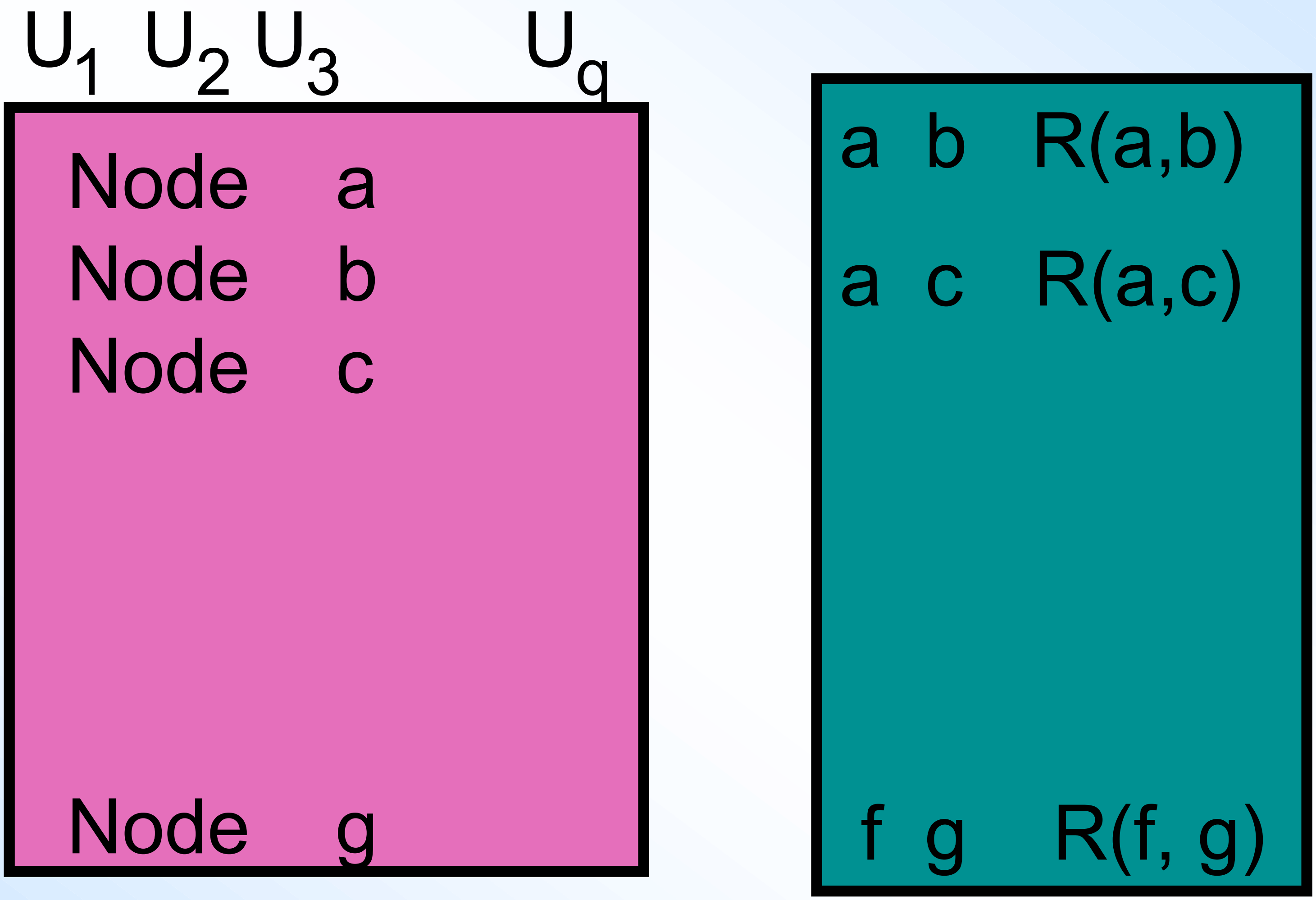


Table #1

Table #2

In the following we shall begin to describe techniques that can be used to analyze, investigate and question networks with vector-valued notes. Here we shall be using flexible/fuzzy queering techniques

First Attribute in Database

- U_1 is country of residence
- Its Domain Y_1 is the set of countries
- Communal vocabulary associated with attribute
- Some terms: Middle East, North America, South America, Southeast Asia, mountainous country, Spanish speaking, "Oil producing"
- Each term in vocabulary defined as subsets of Y_1.

Second Attribute in Database

- U_2 is Age
- Its Domain Y_2 is the set of non-negative integers
- Communal vocabulary associated with attribute
- Some terms: young, old, teenager, senior
- Each term in vocabulary defined as subsets of Y_1.

Retrieve the young people with a path to x_j

$$Ans(x_i) = Young(x_i) \wedge R^n(x_i, x_j)$$

What is the maximal strength of a path connecting node x_j to a person residing in South America？

Letting SA indicate the subset of Y_1 corresponding to South America.

Using this we can obtain as the answer to our question

$$\underset{i, i \neq j}{\mathrm{Max}}[SA(U_1(x_i)) \wedge R^n(x_i, x_j)]$$

What is the maximal strength of a path connecting node x_j to a young person residing in South America ?

$$\underset{i, i \neq j}{\mathrm{Max}}[SA(U_1(x_i)) \wedge Young(U_2(x_i)) \wedge R^n(x_i, x_j)]$$

How true is it that x_j has a *strong* connection to a Young South American?

.

$$\max_{i,i\neq j}[SA(U_1(x_i)) \wedge Young(U_2(x_i)) \wedge Strong(R^n(x_i, x_j))]$$

How true is it that all people in South America have a strong connection with each other ?

$$\min_{i,j}[\overline{(SA(U_1(x_i)) \wedge SA(U_1(x_j)))} \vee Strong(R^n(x_i, x_j))]$$

References

•Yager, R. R., Intelligent social network analysis using granular computing, International Journal of Intelligent Systems 23, 1196-1219, 2008.

•Yager, R. R., "Concept representation and database structures in fuzzy social relational networks," IEEE Transactions on Systems, Man and Cybernetics: Part A 40, 413-419, 2010.

Social Information Sharing Using Fuzzy Methods

Linguistic Summaries

Knowledge Discovery

- Knowledge discovery (mining) in social networks important subject of interest

- Supplies meta-knowledge about objects in a network

- Expressible in manner human can easily understand

- Useful for predicative purposes

- Supplies information for models

Some Key Issues in Data Discovery

- **Representation of Knowledge**

  F ormal language to represent information discovered.

- **Validation**

  M ethod for testing the validity of conjectured observations

- **Measure of Value of Information**

- **Focusing and Conjecturing**

  Tools to help focus on what to look for

Types of Information Discovered

- Summaries
- Relationships and rules
- Dependencies
- Typical Values
- Atypical Values

Why Summarize?

1. To grasp the meaning of data

2. To communicate observations to other people.

3. As a starting point for the ability to make inferences from data.

INFORMATION PRESENTATION

Raw Data

Undigested

Hard to Comprehend

Mean

To Terse

Needs Numbers

Linguistic Summaries

Linguistic Summaries

- Examples of linguistic summaries:

  Few people in the network are **old**

  Most friends of Jim in the Network base are **wealthy**

  Most of Johns friends are highly educated

- User friendly
- Generalizes probabilistic statements

General Structure of Linguistic Summary

Q **A** are **B**

- Q: Quantity in agreement

  Linguistic Quantifier

  Fuzzy subset

B: Linguistic Summarizer

Fuzzy subset over domain of attribute

A: Subpopulation Selector

Fuzzy subset over domain of network elements

Measure of Validity of Summary

Associate a truth value τ with a linguistic summary based on its compatibility with Network summary

Q A are B

$$1.\ r = \frac{\sum_{x \in Net} A(X)B(x)}{\sum_{x \in Net} A(X)}$$

$$2.\ \tau = Q(r)$$

Measure of Informativeness of Summary

Consider two summaries:

Most tall people are **very fat**

Some tall people are **fat**

First is more informative

Consider summary Q A are B

$$I = Sp(B)\ (1 - Sp(A))\ Sp(Q)\ \tau$$

Information is increased if:

1. If B is narrow
2. If A is wide
3. Q is large
4. τ is large

Conjecturing

- Linguistic summaries provide a methodology for representing and validating a type of meta knowledge contained in a social network.

- Many possible summaries can be conjectured !

- How do we select which ones to test

  Human driven

  Data driven

Template Based Methods

- Summary expressed in terms humans use to understand and discuss attributes

- For given attribute provide a partitioning of the domain in linguistic values associated with the attribute

  Height: {short, average, tall}

- Provide class of quantifiers

- Test Summaries of form

  Q_i A_j are B_k

Data Driven Approaches

- Use data to suggest summaries
- Clustering methods
- Mountain Method
- Centers of clusters provide nucleus for conjectured summaries

Social Information Sharing Using Fuzzy Methods

Using Fuzzy Sets to Model Information Provided by Social Tagging

Joint work with Marek Z. Reformat

Tagging

Tags are labels used by users to describe items/resources, they represent users' understanding and perception of items

Items/resources are described by anyone who "sees" them and wants to provide their description and/or comments

Method of sharing Information in social websites

Tag Clouds

Many tags are used to annotate a single resource, and multiplicity of those tags can vary - a graphical representation of resource annotation is called a tag-cloud

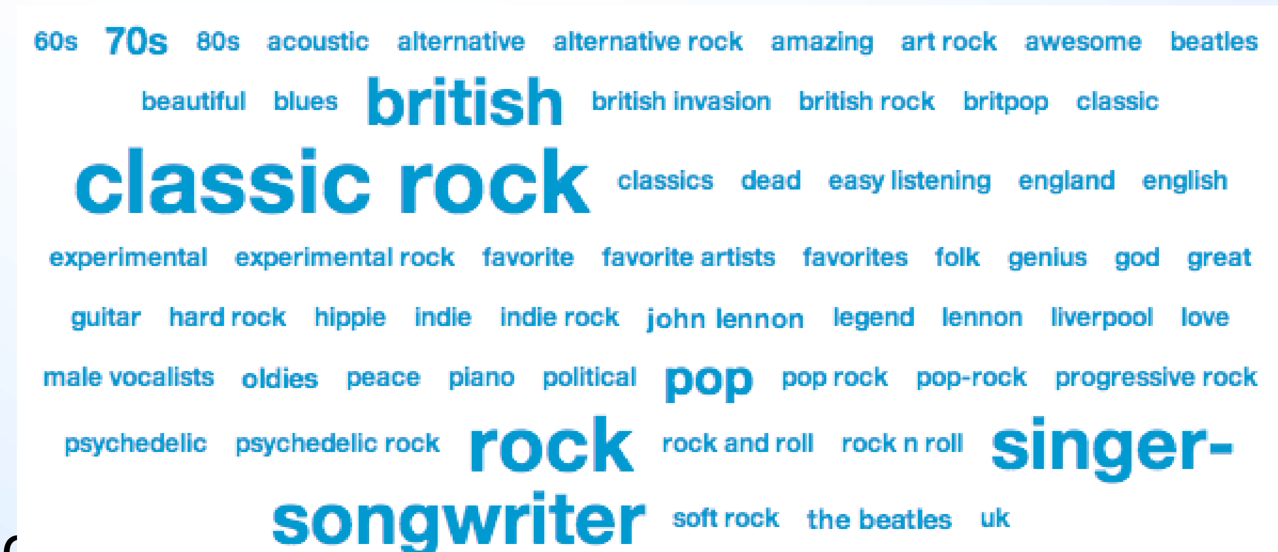
Tag clouds Example

1001 20th century american anthology artificial intelligence **asimov** classic classic science fiction classics collection fantasy fiction future literature made into movie movie novel own paperback read robotics **robots** **sci-fi** **science fiction** series **sf** sff **short stories** speculative fiction unread

Tags for the resource:

I, Robot by Isaac Assimov (librarything.com)

Tag Cloud Example



tags for the resource:

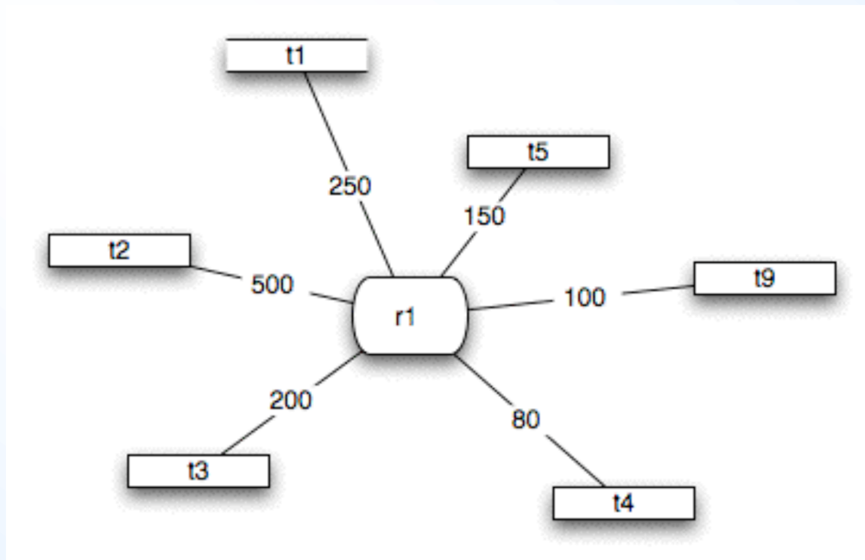
John Lennon (last.fm)

tag clouds

tag cloud for resource *r1*

tag clouds

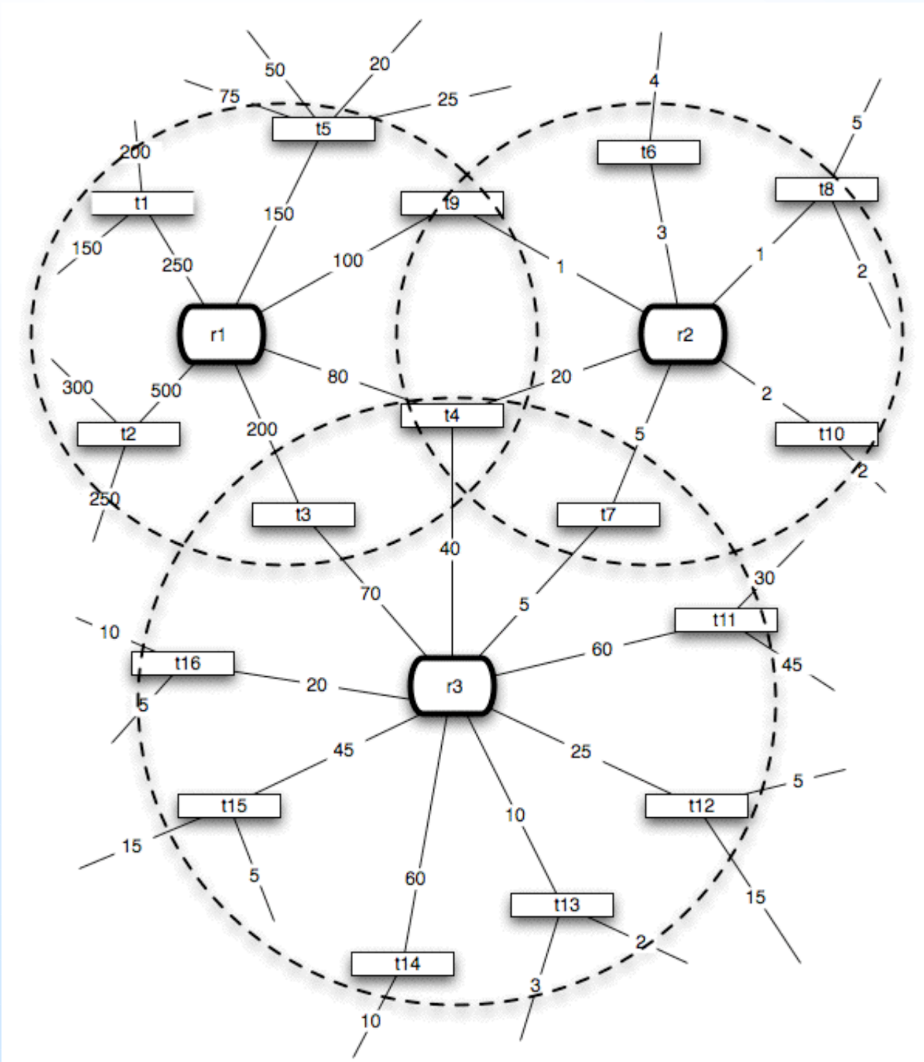
graph representation for resource *r1*



strength of a connection: $occur(t_i, r_j)$

tag clouds

“bigger picture”



Tag clouds “as” fuzzy sets

The size of each tag (its occurrence value) expresses a degree

to which this tag describes or represents the resource

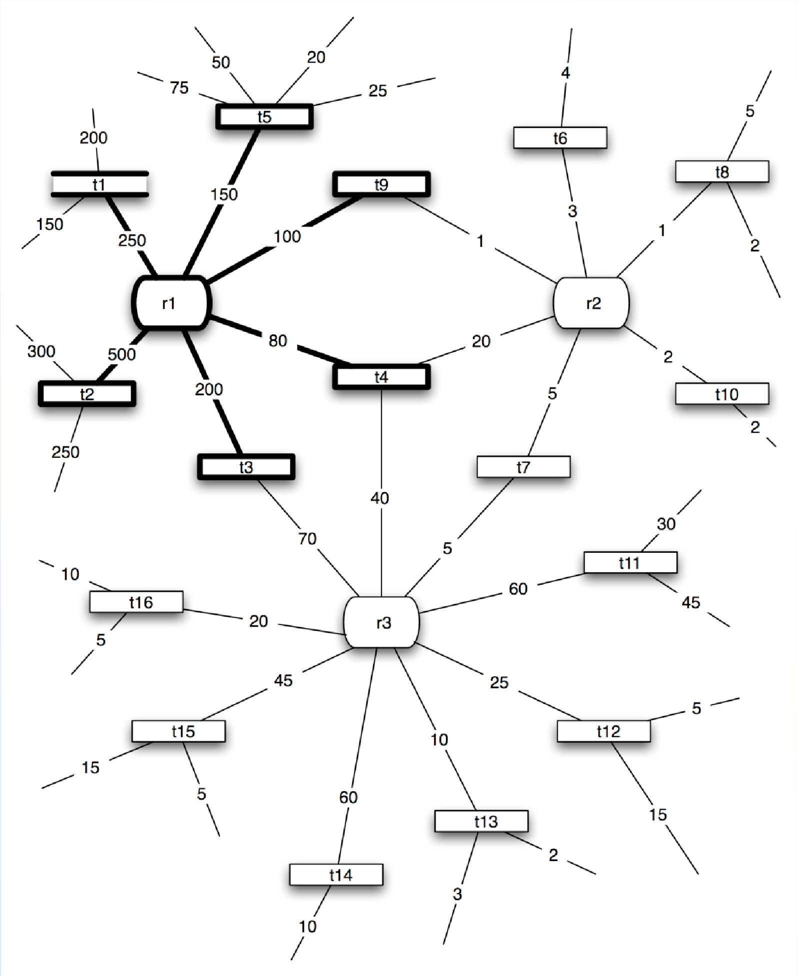
Membership Grade in Fuzzy Set

Tag clouds "as" fuzzy sets

60s 70s 80s acoustic alternative alternative rock amazing art rock awesome beatles
beautiful blues british british invasion british rock britpop classic
classic rock classics dead easy listening england english
experimental experimental rock favorite favorite artists favorites folk genius god great
guitar hard rock hippie indie indie rock john lennon legend lennon liverpool love
male vocalists oldies peace piano political pop pop rock pop-rock progressive rock
psychedelic psychedelic rock rock rock and roll rock n roll singer-
songwriter soft rock the beatles uk

tag_____
John Lennon (last.fm)

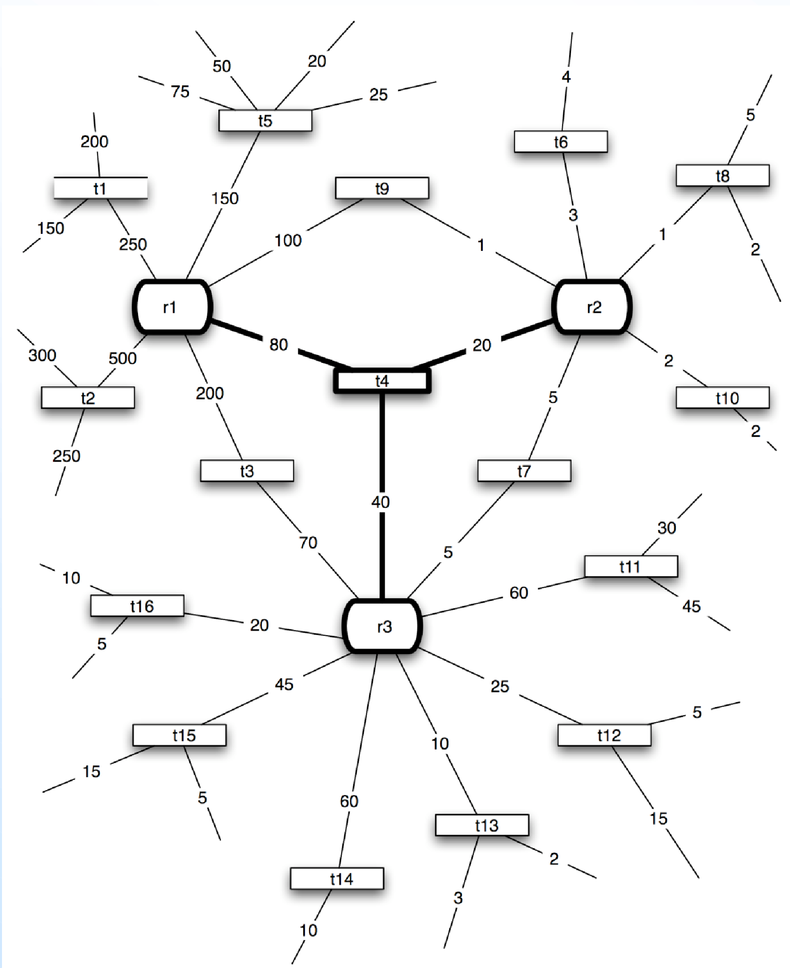
Tag-based Fuzzy sets



Fuzzy set representation of resource

$$\Phi_r(r_j) = \{\frac{\mu_j^r(t_1)}{t_1}, \frac{\mu_j^r(t_2)}{t_2}, \ldots, \frac{\mu_j^r(t_m)}{t_m}\}$$

Tag-based fuzzy sets



fuzzy set representation of tag

$$\Phi_r(t_i) = \{\frac{\mu_i^t(r_1)}{r_1}, \frac{\mu_i^t(r_2)}{r_2}, \ldots, \frac{\mu_i^t(r_n)}{r_n}\}$$

tag-based fuzzy sets

many possible applications of
tag-based fuzzy sets
one of them – item selection

item selection

task:

to identify a resource with the highest degree of relationship to a given keyword (tag)

conclusions

- building fuzzy sets based on data and information provided by Internet users
- utilization of those fuzzy sets for building systems supporting user activities in social networks and the Internet

REFERENCES

[1]. Yager, R. R. and Reformat, M., "Tagging and Fuzzy Sets," Intelligent Systems - From Theory to Practice, V. Sgurev, M. Hadjiski and J. Kacprzyk (Eds.), Springer Verlag, Heidelberg, 1-18, 2010.

[2]. Yager, R. R. and Reformat, M. Z., "Using fuzzy sets to model information provided by social tagging," Proceedings of the Fuzz-IEEE at the World Congress on Computational Intelligence WCCI 2010, Barcelona, 3258-3265, 2010.

[3]. Reformat, M. Z. and Yager, R. R., "Tag-based fuzzy sets for criteria evaluation in on-line selection processes," Journal of Ambient Intelligence and Humanized Computing 2, 35-51, 2011.

Structure and Dynamics of Networks

Some Useful Concepts

1. Degree of Node (number of links) $d_j = \sum_{j \neq i} R(x_i, x_j)$

2. Coordination Coefficient (average links per node) $z = \frac{1}{n} \sum_{j=1}^{n} d_j$

3. Proportion of nodes with k links (Probability) p_k

4. Distance between two nodes: Geo(x, y)

   Minimum k s.t. $R^k(x, y)=1$

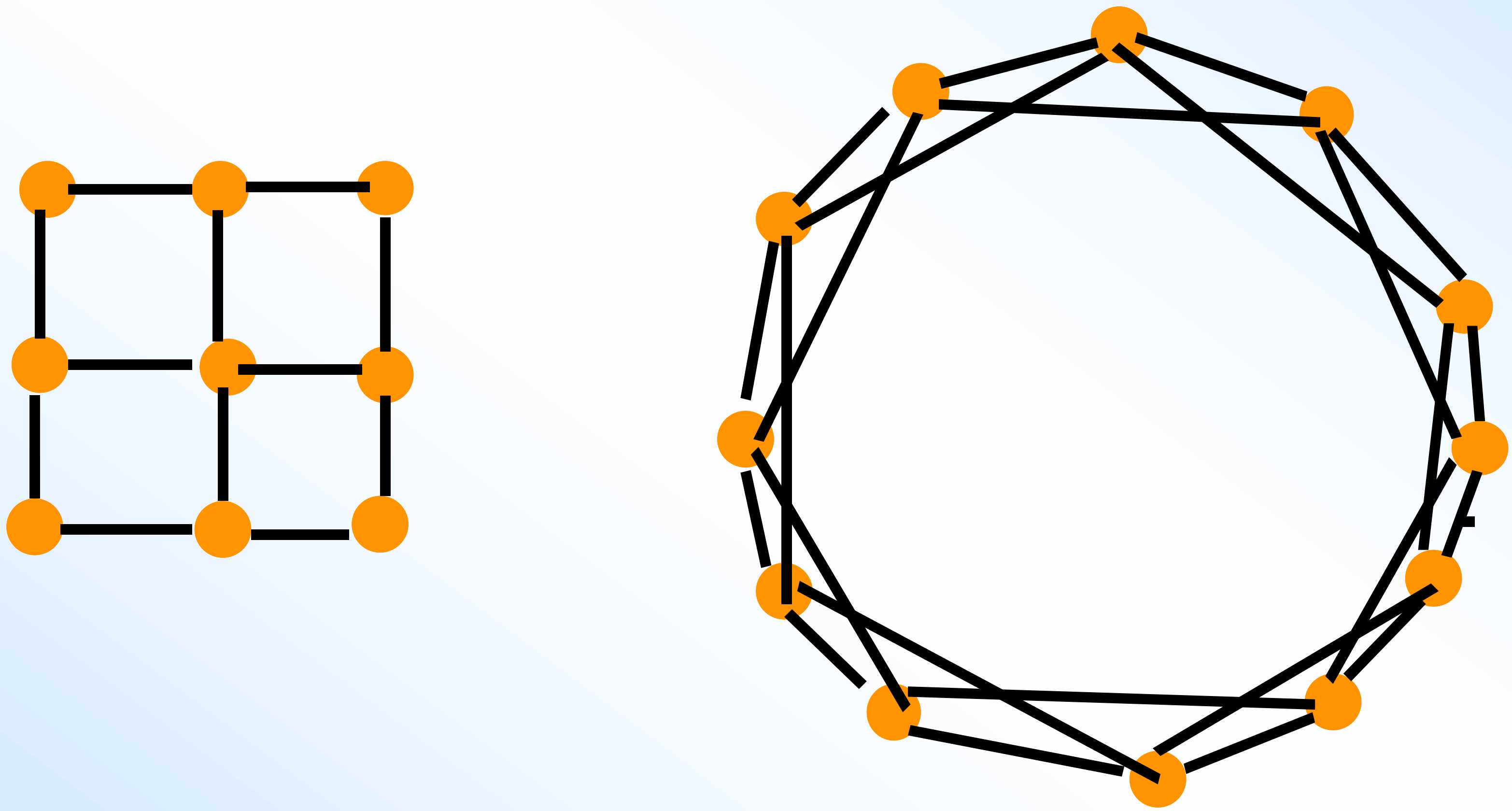
5. L average Geo

6. Network Diameter; Largest Geo

Some Notable Network Structures

1. Lattice Structured
2. Pure Random
3. Scale Free
4. Small World

Lattice Structured Networks



Pure Random Networks

Erdös-Rényi

If we add new node its probability of

connecting to an existing node is p

Properties of Random Networks

Average links per node $z = p\,(n-1)$

$$p_k = e^{-z}\frac{z^k}{k!} \quad \text{(for large n)} \quad \text{Poisson Distribution}$$

Average distance between nodes: $L = 1 + \dfrac{\ln(n)}{\ln(z) + \ln(z-1)}$

Scale Free Networks

Generative Paradigm

<u>Preferential Attachment</u>

Probability that a new node x_{n+1} connects with a existing node x_j is proportional to the number of links d_j that x_j has

$$\mathrm{Prob}(R(x_{n+1}, x_j)) = 1) \propto \frac{d_j}{\sum_{i=1}^{n} d_i}$$

Features of Scale Free Networks

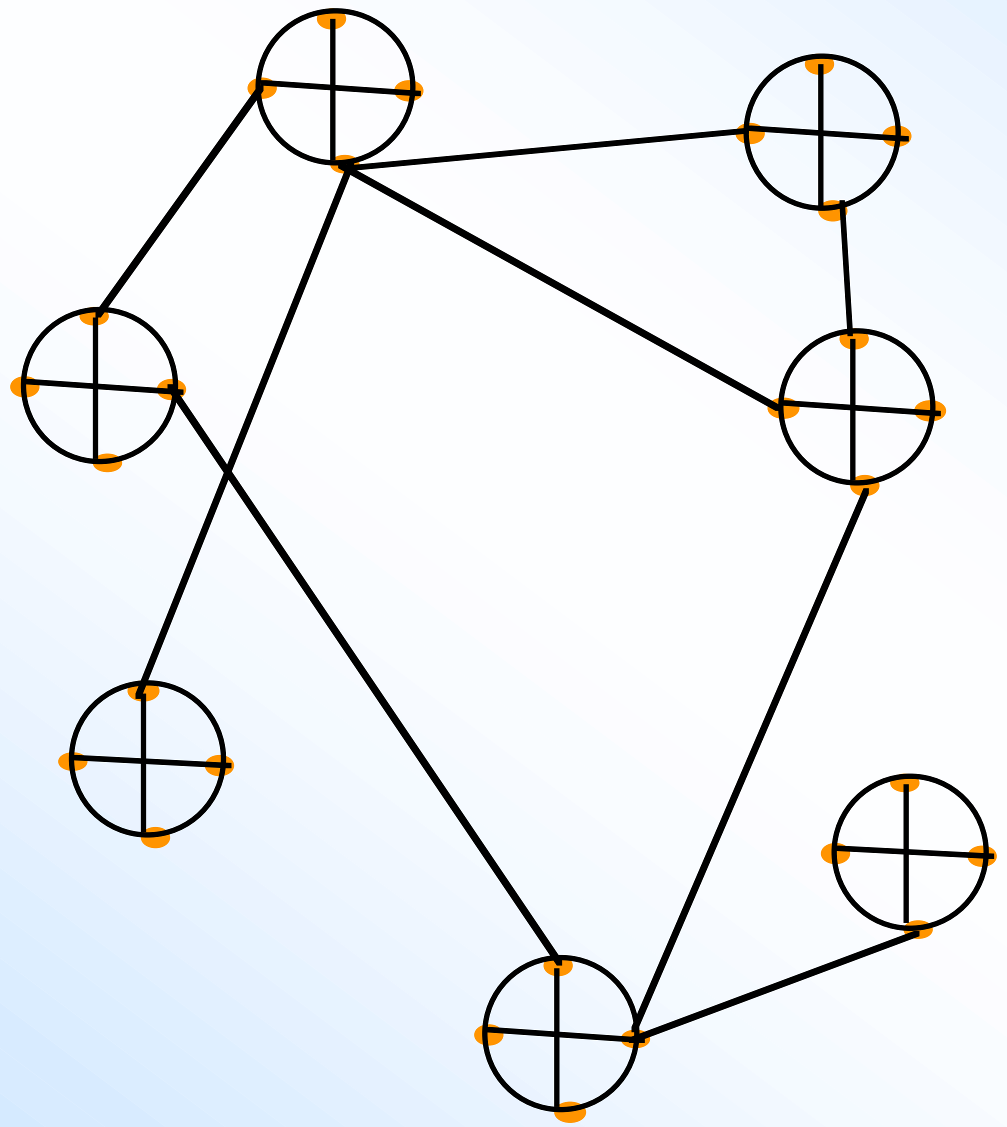
- Proportional of nodes having k links

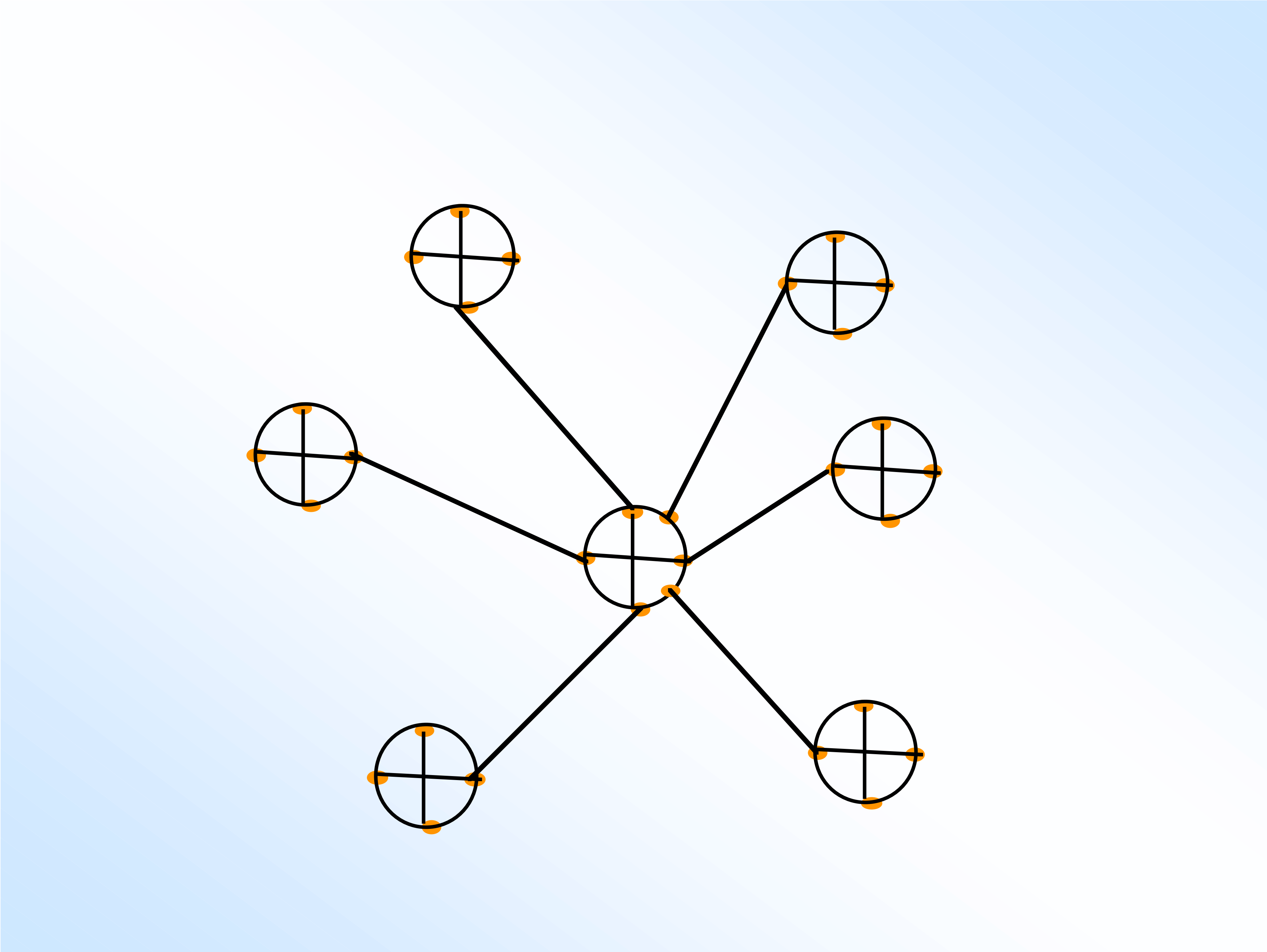
$$p_k = \frac{1}{k^{\lambda}} \qquad \lambda \approx 3$$

- Few nodes with many links (Hubs)
- Most nodes with few links
- Many networks are of this type

    Web Pages on Internet

Small Worlds Network





Features of Small Worlds Networks

Primary Features

Highly Clustered Small Neighborhoods

Small Shortest Distance Between Nodes

Secondary Feature

Abundance of Hubs

Fat Tailed/Scale free like

Clustering Coefficient

Proportion of pairs of neighbors of a node that are also neighbors of each other.

Quantifies how close neighbors are to being a cliques

Neighborhood of Node x_i

- Degree x_j is neighbor of x_i

$$\mathrm{Neigh}_i(x_j) = R(x_i, x_j)$$

- Neighborhood of x_i

$$N_i = \{x_j \mid \mathrm{Neigh}_i(x_j) = 1\}$$

- Size of Neighborhood of x_i

$$d_i = \mathrm{Card}(N_i)$$

Clustering Coefficient of Node x_i

$$C_i = \frac{\text{pairs of neighbors connected}}{\text{possible connections between neighbors}}$$

$$C_i = \frac{2}{d_i\,(d_i - 1)} \sum_{k=j+1}^{n} \sum_{\substack{j=1 \\ i \neq i}}^{n} \mathrm{Neigh}_i(k)\,\mathrm{Neigh}_i(j)\,\mathrm{Neigh}_j(k)$$

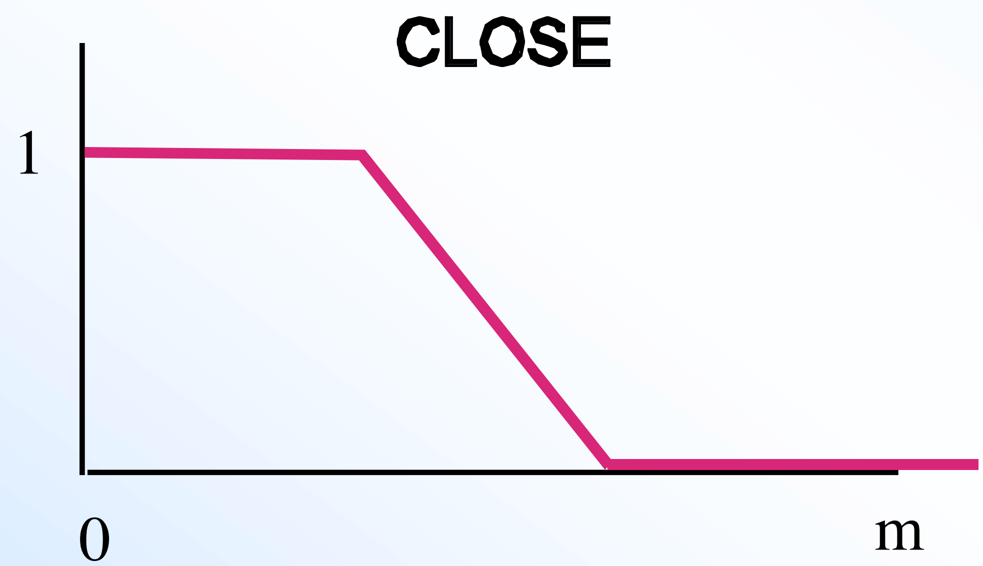
Clustering Coefficient of Whole Network

$$\overline{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$$

Clustering Coefficient

Softer-Fuzzy Definition

Proportion of pairs of close neighbors of a node that are close neighbors of each other



Let Q(m) be degree m links satisfies close

Degree node x_j is close neighbor of x_i

$$QNeigh_i(x_j) = Max_m[Q(m) \wedge R^m(x_i, x_j)]$$

Soft Clustering Coefficient of Node x_i

$$C_I = \frac{\sum_{k=j+1}^{n} \sum_{\substack{j=1 \\ i \neq i}}^{n} QNeigh_i(k)\, QNeigh_i(j)\, QNeigh_j(k)}{\sum_{k=j+1}^{n} \sum_{\substack{j=1 \\ i \neq i}}^{n} QNeigh_i(k)\, QNeigh_i(j)}$$

Short Distance Between Nodes

Communal Vocabulary has Short Distance as a Fuzzy Set SD

SD(k) degree k links is short distance

Satisfaction of Criteria there is a Short Distance between x_j and x_i

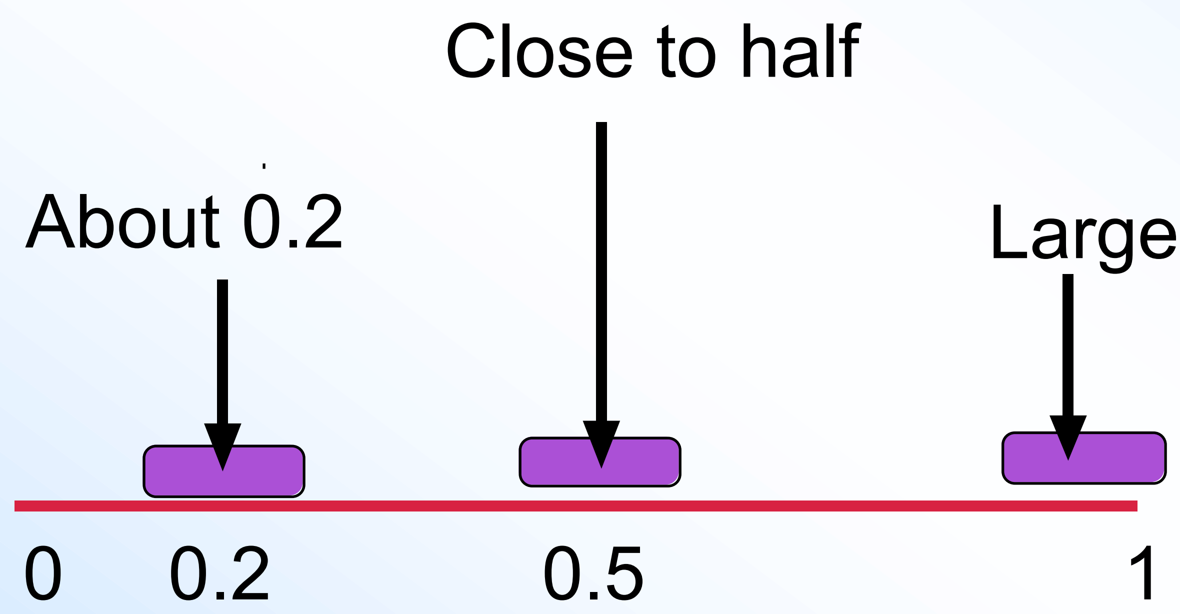
$$Sat(x_i, x_j) = Max_k [SD(k) \wedge R^k(x_i, x_j)]$$

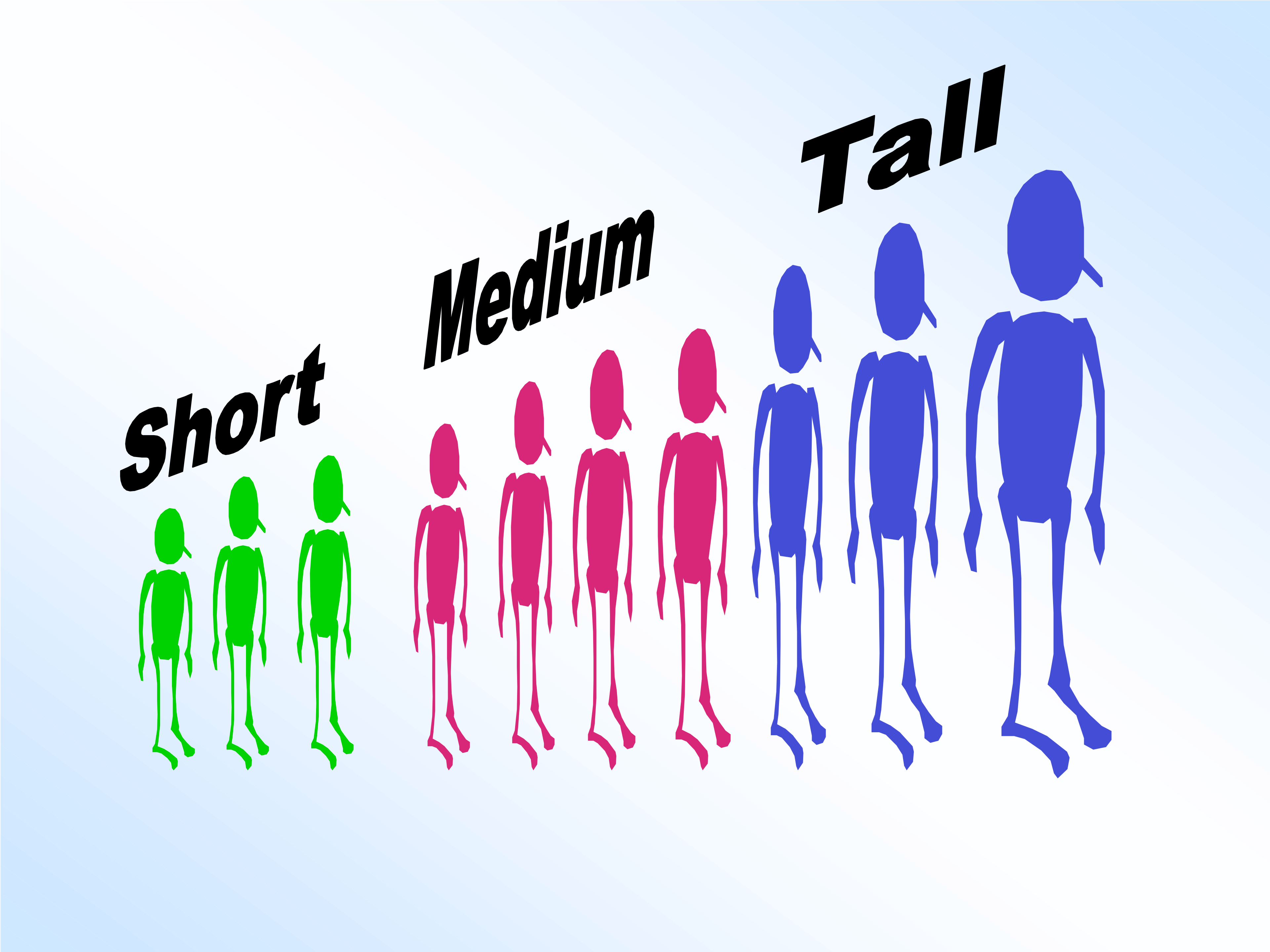
Overall Network Satisfaction

Can Use Sat(x_i, x_j) in Various Ways

- All pairs are SD
- Almost all pairs are SD
- Most pairs are SD
- Pair Average

Granular Probabilities




Short
Medium
Tall

tagging

items/resources are described by anyone who “sees” them and wants to provide their description and/or comments

overview

- introduction
- tag clouds
- tag clouds and fuzzy sets
- fuzzy set for item selection
  - fuzzy presets
  - construction of fuzzy sets
- conclusions

introduction

issue:

finding relevant items on the Internet and application of fuzzy technology for that purpose

introduction

inspiration and opportunity:

web 2.0 and social networks

(users’ involvement in building web contents)

introduction

inspiration and opportunity (examples):

del.icio.us, flickr.com, citeulike.org, libraryThing.com, last.fm, amazon.com

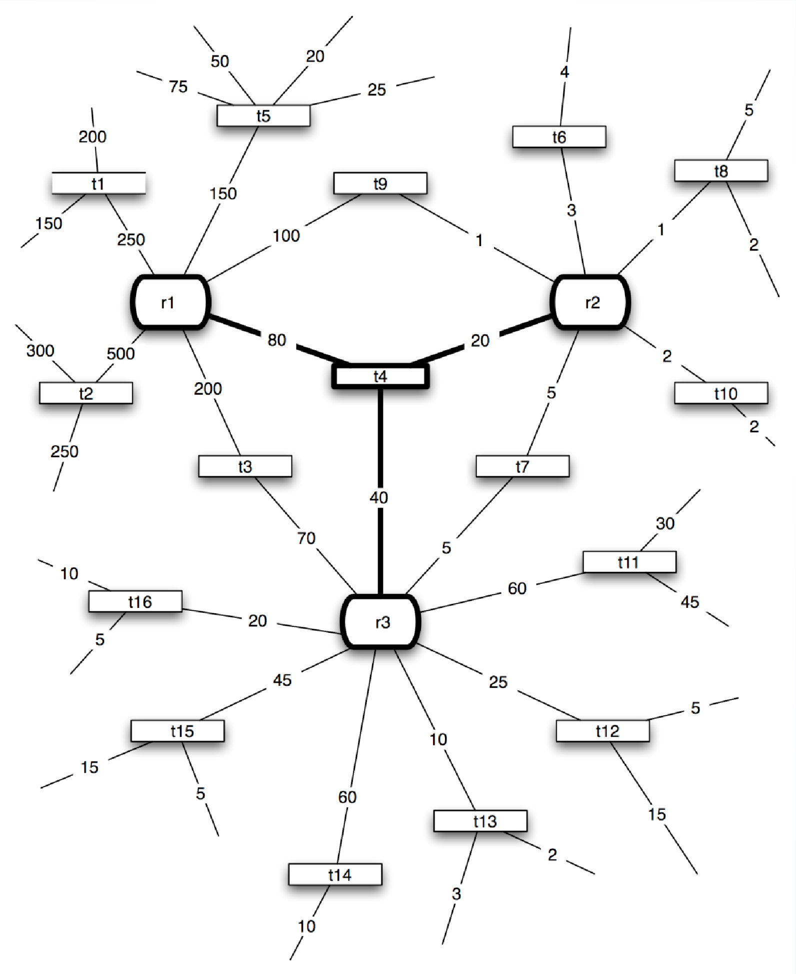
introduction

idea:

construction of fuzzy sets based on the content provided by users and utilization of those sets to support search activities

item selection

tag-based fuzzy set

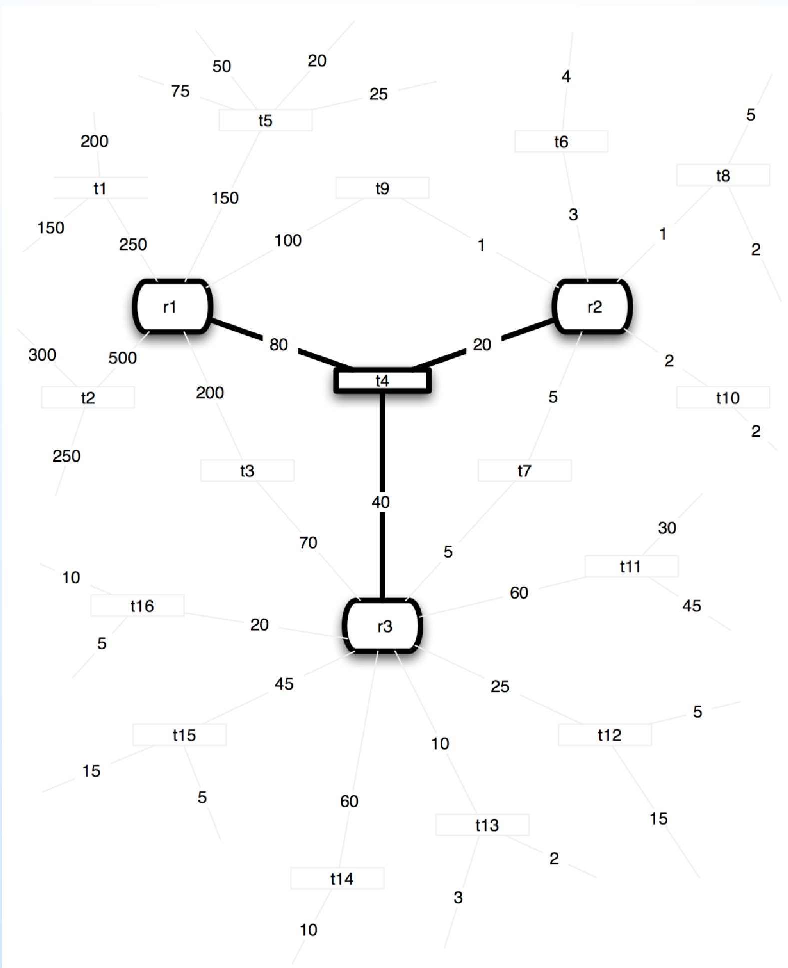


how to estimate the strength (degree) of relationship

$$\Phi_r(t_i) = \{\frac{\mu_i^t(r_1)}{r_1}, \frac{\mu_i^t(r_2)}{r_2}, \ldots, \frac{\mu_i^t(r_n)}{r_n}\}$$

item selection

tag-based fuzzy set

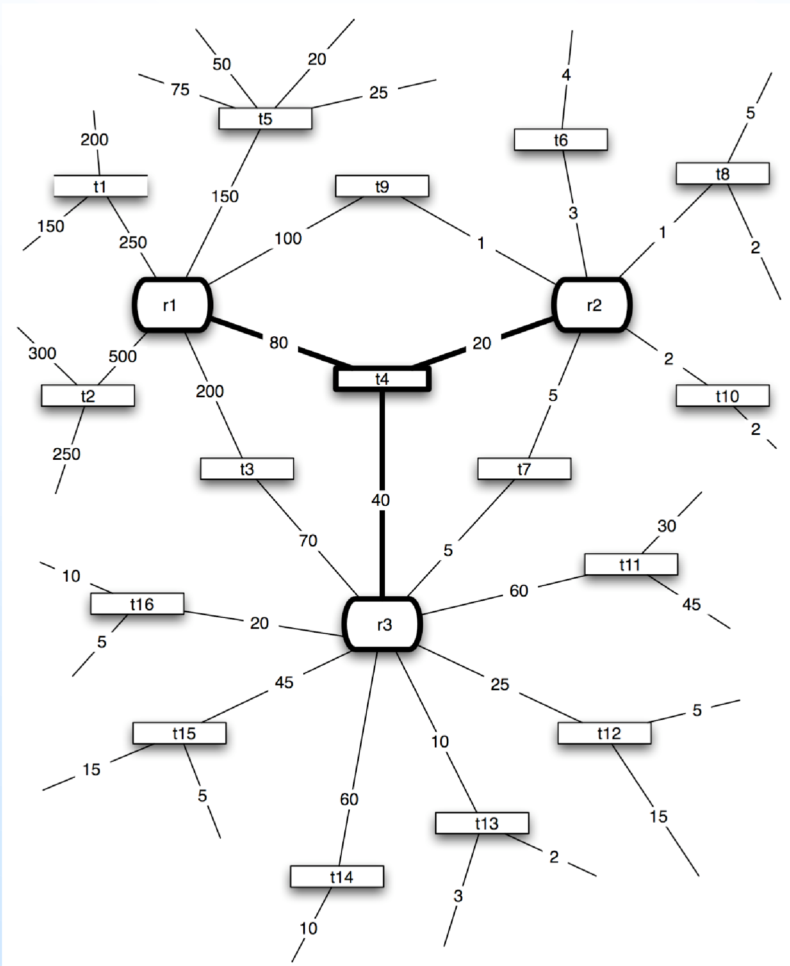


how to estimate the strength (degree) of relationship

$$\Phi_r(t_i) = \{\frac{\mu_i^t(r_1)}{r_1}, \frac{\mu_i^t(r_2)}{r_2}, \ldots, \frac{\mu_i^t(r_n)}{r_n}\}$$

item selection

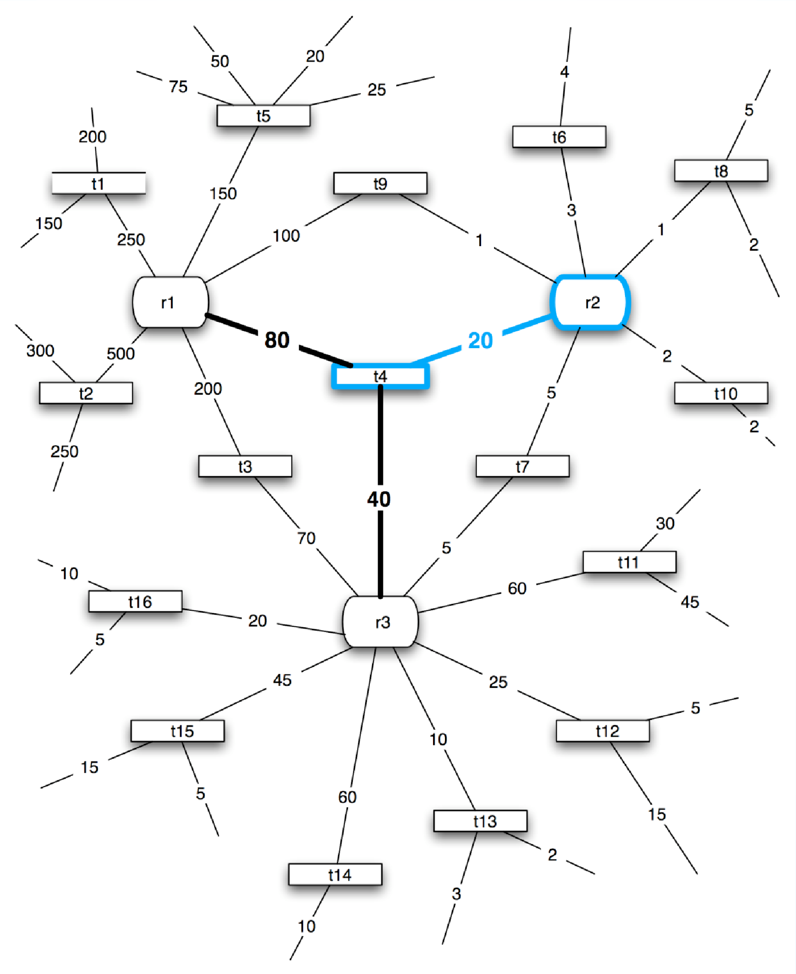
tag-based fuzzy set



how to estimate the strength (degree) of relationship

$$\Phi_r(t_i) = \{\frac{\mu_i^t(r_1)}{r_1}, \frac{\mu_i^t(r_2)}{r_2}, \ldots, \frac{\mu_i^t(r_n)}{r_n}\}$$

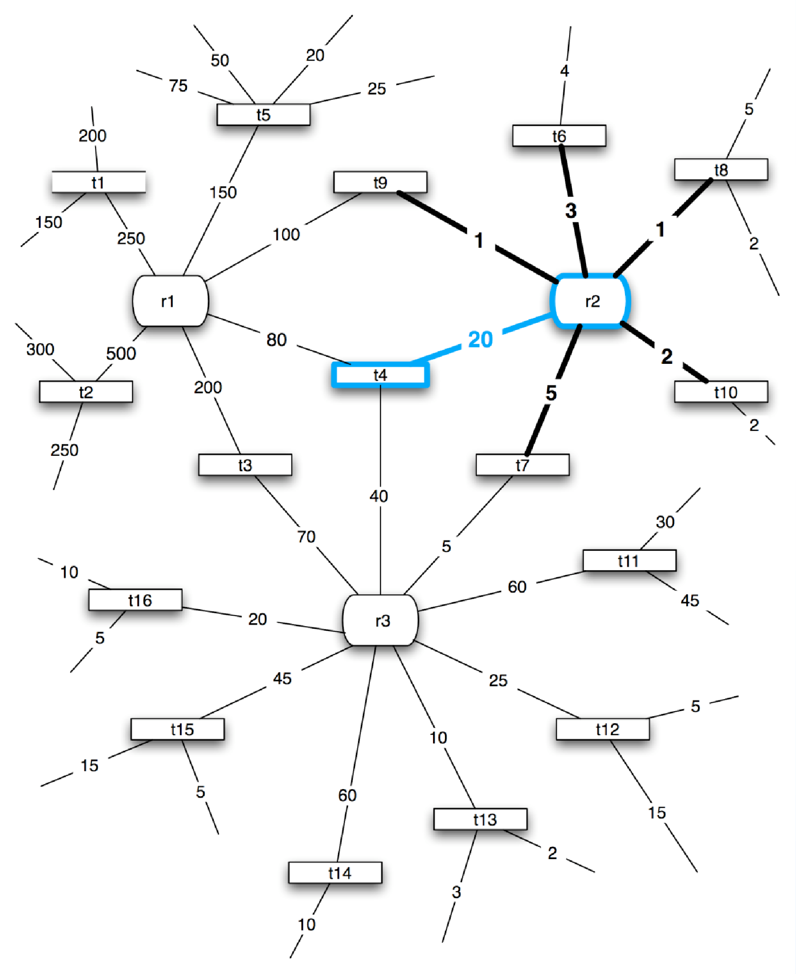
tag-based fuzzy set for item selection



aspect: local (resource)

how the tag-resource relation is doing when compared with other relations the tag has with other resources

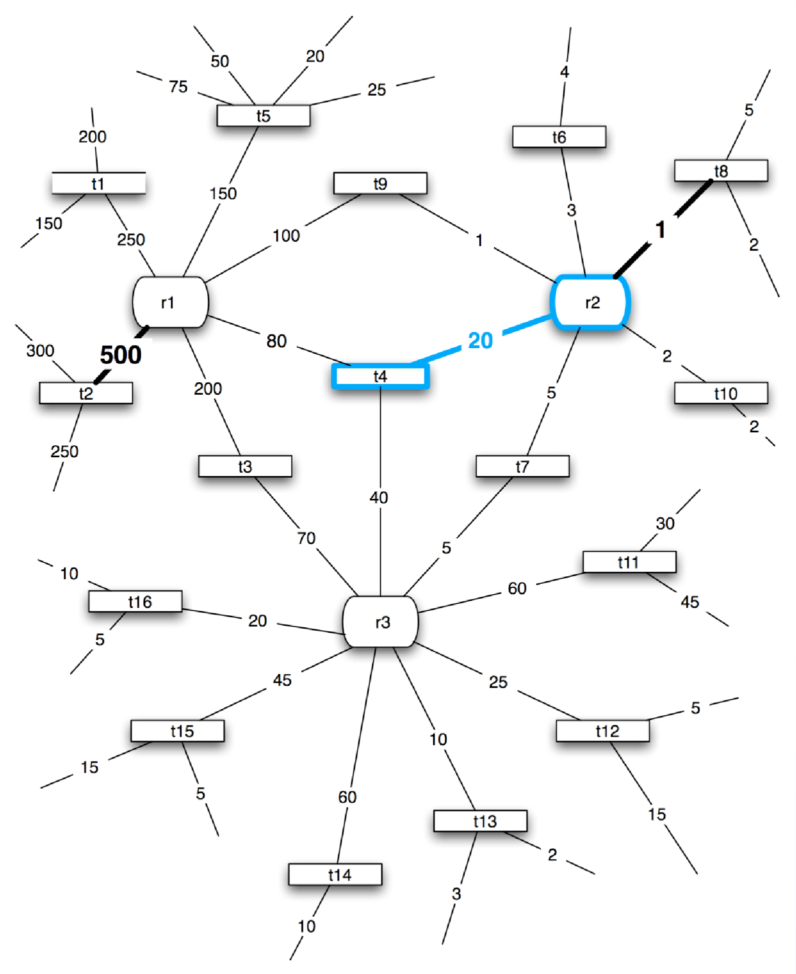
tag-based fuzzy set for item selection



aspect: local (tag)

how the tag-resource relation is doing when compared with other relations the resource has with other tags

tag-based fuzzy set for item selection



aspect: global

how the tag-resource relation is doing when compared with the weakest and best relations of the whole network

tag-based fuzzy set for item selection

three aspects -> three fuzzy pre-sets

- local (resource-based)
- local (tag-based)
- global

tag-based fuzzy set for item selection

two issues

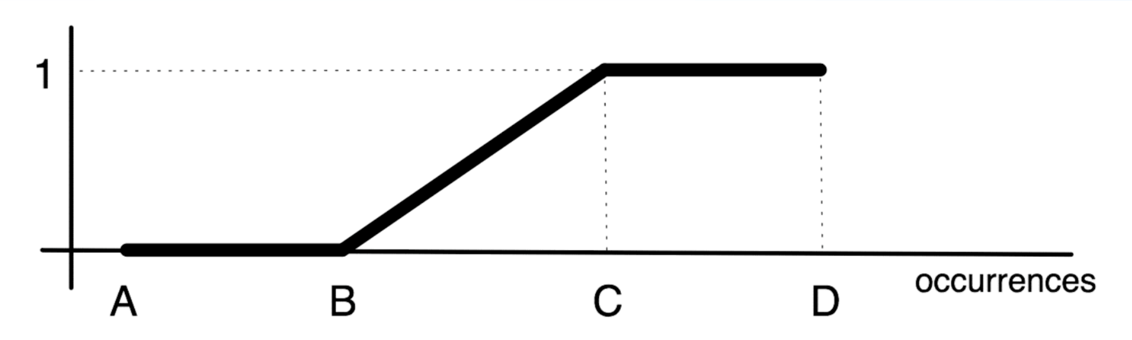
- how to determine values representing those aspects
- how to combine them

tag-based fuzzy set

value determination

a piece-wise linear function built for each aspect based on occurrences *occur*(t_i, r_j) valid for this aspect

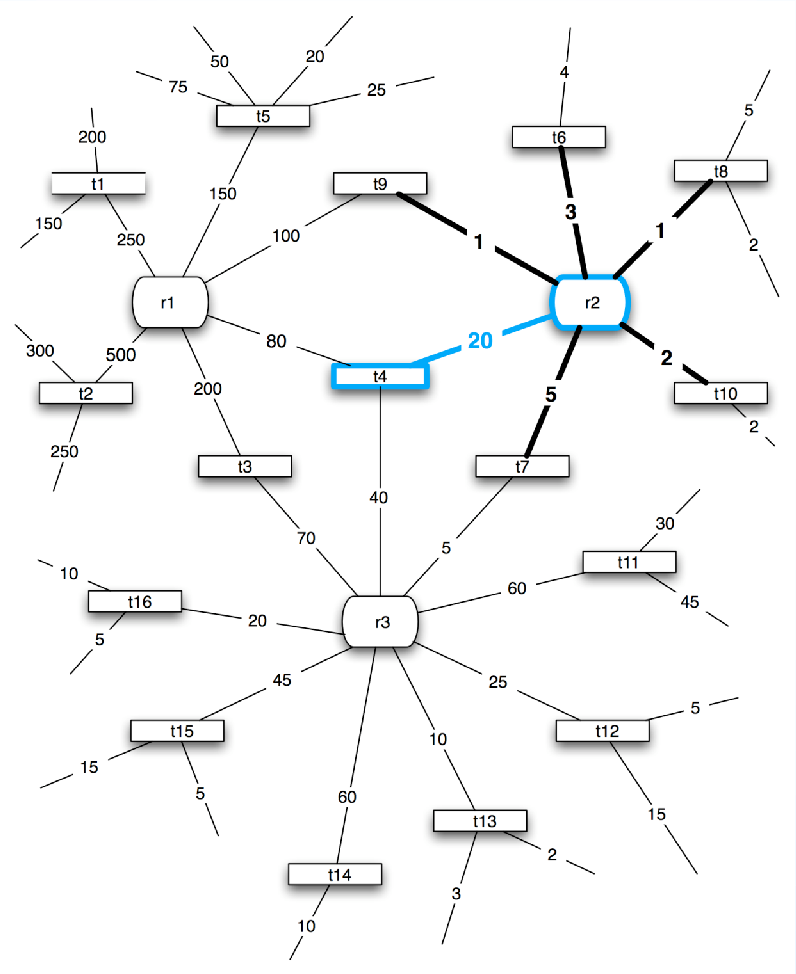
tag-based fuzzy set
value determination



min & max of occurrences valid for a given aspect
(A – min, B – 0.2*D, C – 0.8*D, D – max)

tag-based fuzzy set

value determination



A = 1
B = 4
C = 16
D = 20

tag-based fuzzy set

aggregation of fuzzy presets

three fuzzy presets:

- local (resource) $lr_{i,j}$
- local (tag) $lt_{i,j}$
- global: $g_{i,j}$

tag-based fuzzy set

aggregation of fuzzy presets

fuzzy set:

an aggregation of linguistic statements representing different types of human ways of determining the strength of relationship

tag-based fuzzy set

global

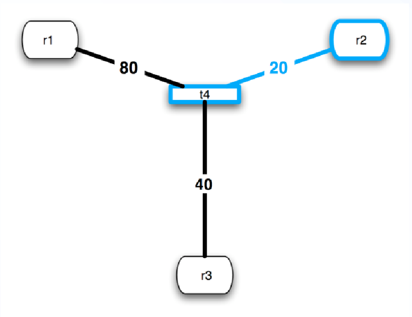
if the occurrence occur(t_i, r_j) of the tag t_i as a label for r_j is globally high then the description strength of t_i should be high

$$g_{i,j}$$

tag-based fuzzy set
local (resource)

if the occurrence occur(t_i, r_j) is globally okay and its value is high when compared with other occurrences involving t_i then the description strength of t_i should be high

$$T(g_{i,j}) \;\; AND \;\; lr_{i,j}$$

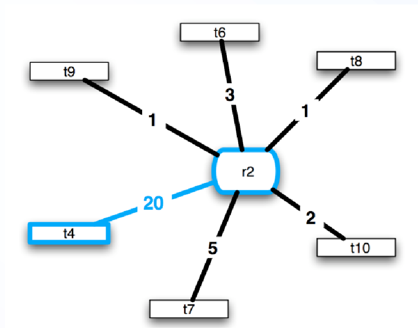


tag-based fuzzy set

local (tag)

if the occurrence occur(t_i, r_j) is globally okay and its value is high when compared with other occurrences involving r_j then the description strength of t_i should be high

$$T(g_{i,j}) \ \ AND \ \ lt_{i,j}$$



tag-based fuzzy set

aggregation of fuzzy presets

$$\begin{aligned}\mu_i^t(r_j) \quad = \quad & g_{i,j} \quad OR \\ & (T(g_{i,j}) \;\; AND \;\; lr_{i,j}) \quad OR \\ & (T(g_{i,j}) \;\; AND \;\; lt_{i,j})\end{aligned}$$

where $T(g_{i,j}) = g_{i,j}^{\alpha} \quad \alpha \in \langle 0,1)$

tag-based fuzzy set

item selection

determining $\mu^t_i(r_j)$ for each pair (t_i, r_j) (where j changes over all resources annotated with t_i)

identifying the one the highest value