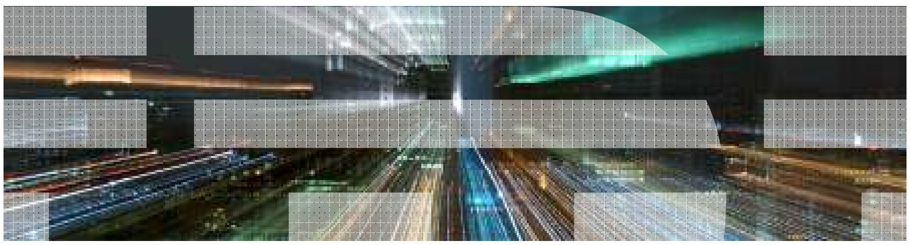
E6885 Network Science Lecture 6:
Network Models -- I

Ching-Yung Lin, Dept. of Electrical Engineering, Columbia University

October 18th, 2010



Course Structure

Class Date	Class Number	Topics Covered
09/13/10	1	Overview – Social, Information, and Cognitive Network Analysis
09/20/10	2	Network Representations and Characteristics
09/27/10	3	Network Partitioning and Clustering
10/04/10	4	Network Visualization
10/11/10	5	Network Sampling and Estimation
10/18/10	6	Network Models -- I
10/25/10	7	Network Models -- II
11/08/10	8	Network Topology Inference -- I
11/15/10	9	Network Topology Inference -- II
11/22/10	10	Dynamic Networks -- I
11/29/10	11	Dynamic Networks -- II
12/06/10	12	Social Influence and Info Diffusion in Networks -- I
12/13/10	13	Social Influence and Info Diffusion in Networks -- II
12/20/10	14	Final Project Presentation

What is a complex network?

- Most real-world networks have complex topological features:
  - Heavy-tail in the degree distribution
  - High clustering coefficient
  - Community structure at many scales
  - Self-similar hierarchical structure

- Simple networks:
  - Typically represented by graphs such as a lattice or a random graph.
  - Topology structure roughly the same in any part of network.
  - Does not posses the above features

- Examples:
  - Social Networks – studied in sociology, public health, commerce, communication.
  - Computer Networks – WWW, security,…
  - Biological Networks – neurons, genes, protein, animals,…
  - Others: sensor network, river network, power lines, …

Measurement of Complexity (I)

- The basic goal of data mining is prediction

- Complexity can be defined as the amount of information required for optimal prediction. (Grassberger, J of Theoretical Physics,1986)

$$C = \min_{f \in M} H[f(X^-)]$$

- f is any predictor that translates the past of the time sequence x^- (or, in other occasions, training set) into an effective state, $s=f(x^-)$, and then make its prediction on the basis of *s*.

Measurement of Complexity (II)

- Grassberger-Crutchfield-Yong Statistical Complexity (J. of Statistical Physica, 2001)
- An effective procedure for finding the minimal maximally predictive model and its states.
- Definition: *causal states* of a process:
  – Two histories x_1^- and x_2^- are equivalent if $\Pr(X^+ \mid x_1^-) = \Pr(X^+ \mid x_2^-)$
  – Write the set of all histories equivalent as $[x^-]$
  – A function which maps each history into its equivalent class: $\varepsilon(x^-) = [x^-]$

$$\Pr[X^+ \mid \varepsilon(x^-)] = \Pr[X^+ \mid x^-]$$

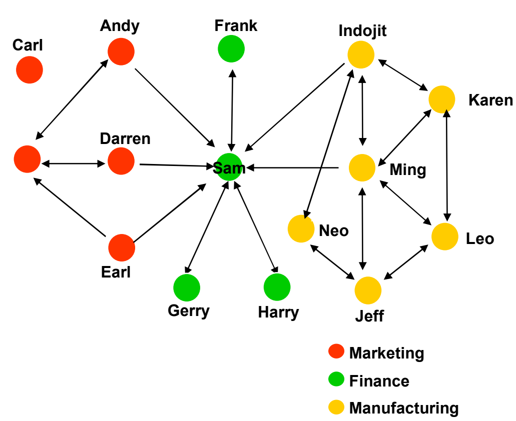
  – Crutchfield and Young proposed to forget particular history and retain only its equivalent class. They call the equivalent classes as the causal states of a process. These are the optimal states.
  – The statistical complexity of a processes is thus the information content of its causal states.
  – It is equal to the shortest description of the past which is relevant to the actual dynamics of the system> E.g., IID: *0*, periodic sequences: *log p*.

Group and Roles

Central people
– Sam. Could be bottleneck or holding group together

Peripheral people
– Earl. Goes to others but no-one goes to him for information. At risk for leaving. Potentially unrealized expertise

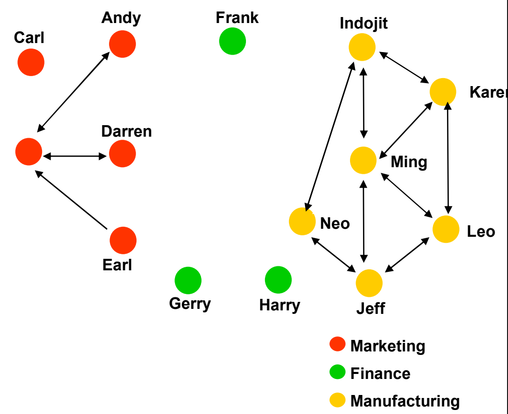
Sub-groups
– Group split by function. Very little information shared across groups



This slide is excerpted from SNA Theory, Concepts and Practice by Dr. T. Mobbs, BCS and Dr. K. Ehrlick, Research

Some Roles are especially critical

What happens if Sam leaves the group through layoffs, job reassignment, attrition, merger, retirement?



This slide is excerpted from SNA Theory, Concepts and Practice by Dr. T. Mobbs, BCS and Dr. K. Ehrlick, Research

Relationships are multi-dimensional and (traditionally) uncovered through network questions

Actions | **Awareness** | **Emotional**

Communication
How often do you communicate with this person?

Innovation
How often do you turn to this person for new ideas

Advice
How often do you seek advice from this person before making an important decision?

Awareness
I am aware of this person's knowledge and skills

Valued Expertise
How likely are you to turn to this person for specialized expertise

Learning
How likely are you to rely on this person for advice on new methods and processes

Trust
I believe there is a high personal cost in seeking advice or support from this person

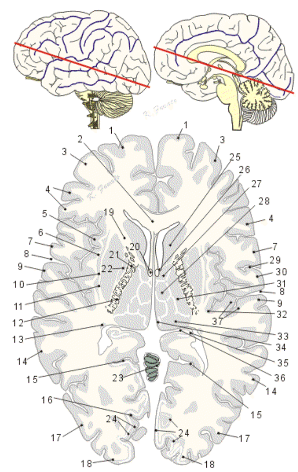
Access
I believe this person will respond to my request in a reasonable and timely manner

Energy
I generally feel energized when I interact with this person

Provided by Drs. Tony Mobbs and Kate Ehrlich, IBM

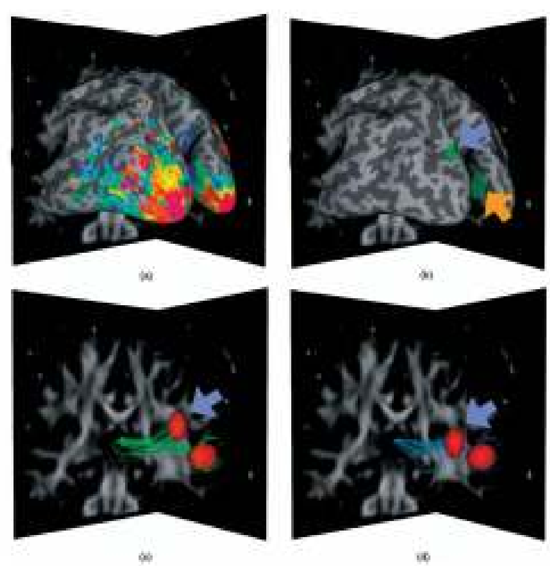
Complex Network in Brain

- Diffusion Tensor Imaging (DTI) provides an important complement to functional magnetic resonance imaging (fMRI).
- fMRI reveals gray matter areas that are metabolically active during performance of a particular behavior or cognitive task.
- It can be considered “modern-day phrenology,” assigning functional roles to parcels of brain tissue with a limited view of the brain’s powerful capacity to function as an interactive network, integrating information across several anatomical sites to produce behavior.
- The combination of fMRI and DTI will provide important insights into these types of neurobehavioral networks by simultaneously revealing active gray matter areas and the white matter pathways that connect them.

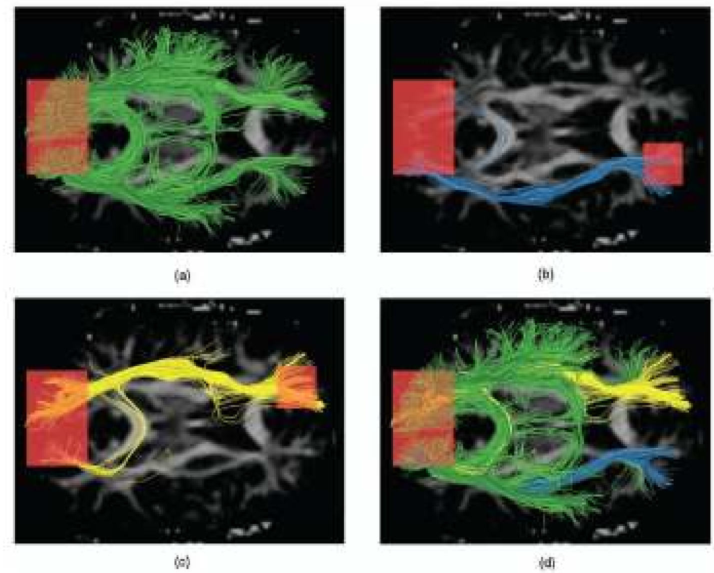


Complex Network in Brain

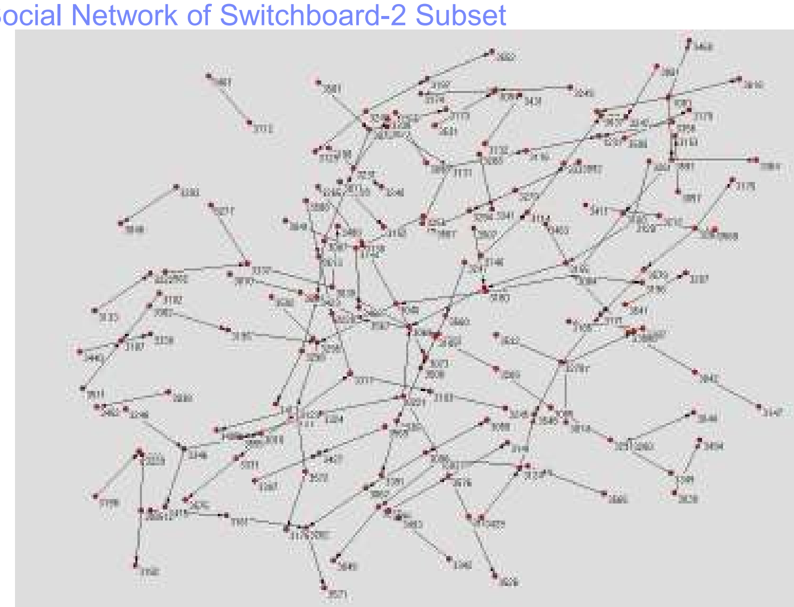
- DTI has been successfully used to describe white matter development in pediatric samples.
- Changes in white matter diffusion properties are consistent across studies, with anisotropy increasing and overall diffusion decreasing with age.
- Diffusion measures in relevant white matter regions correlate with behavioral measures in healthy children and in clinical pediatric samples.



Complex Network in Brain

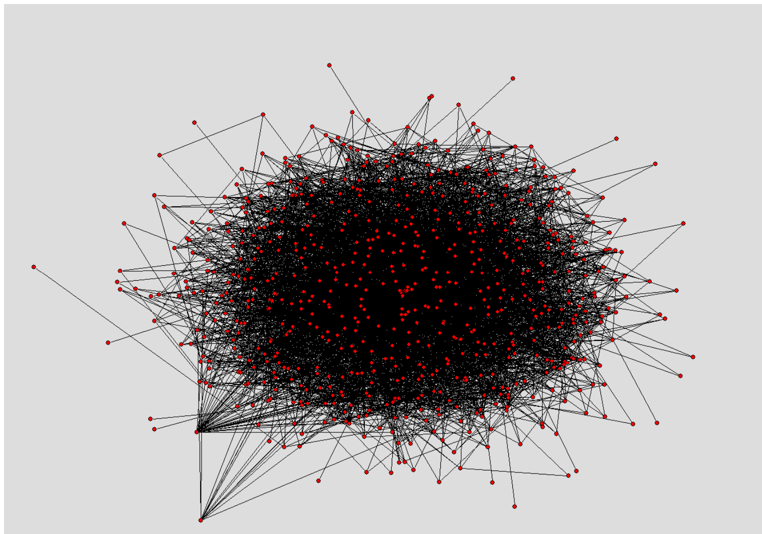


Social Network of Switchboard-2 Subset



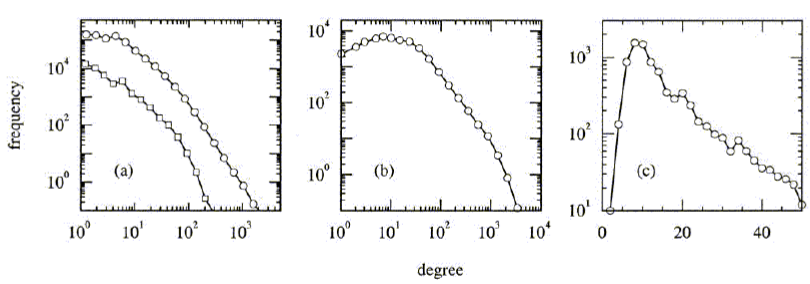
Social Network of Switchboard-2 Dataset

- 679 nodes ➔ edges = 4472



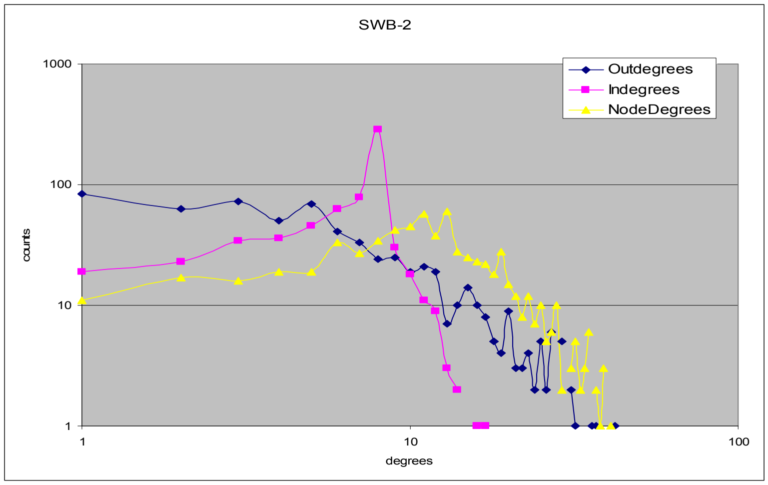
Some examples of Degree Distribution

- (a) scientist collaboration: biologists (circle) physicists (square), (b) collaboration of move actors, (d) network of directors of Fortune 1000 companies



Switchboard-2 Network Degree Distribution

- 679 nodes (actors)
- Out degrees ➔ Normal. In degrees ➔ Abnormal.

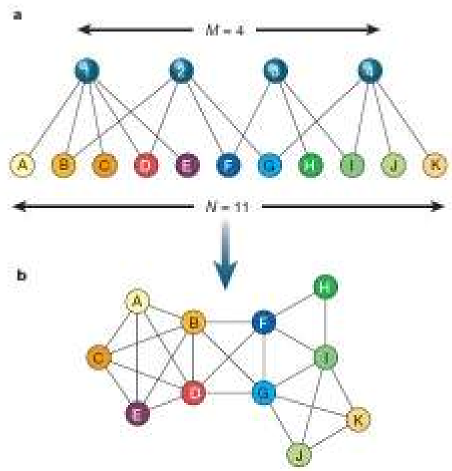


Two Types of Random Graphs

- Directed/Undirected Graphs
- Intrinsic Bipartite or multi-partite graphs ➔ Unipartite Graphs (One-Mode Graphs)
  – Existence of Groups/Communities

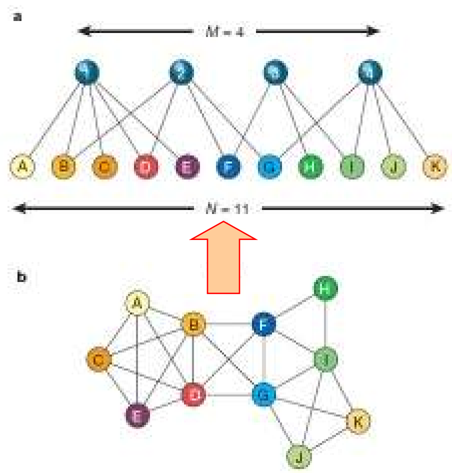
Bipartite Graphs ➔ One-Mode Network

- There exists intrinsic structures:
  - Example: M boards, N directors



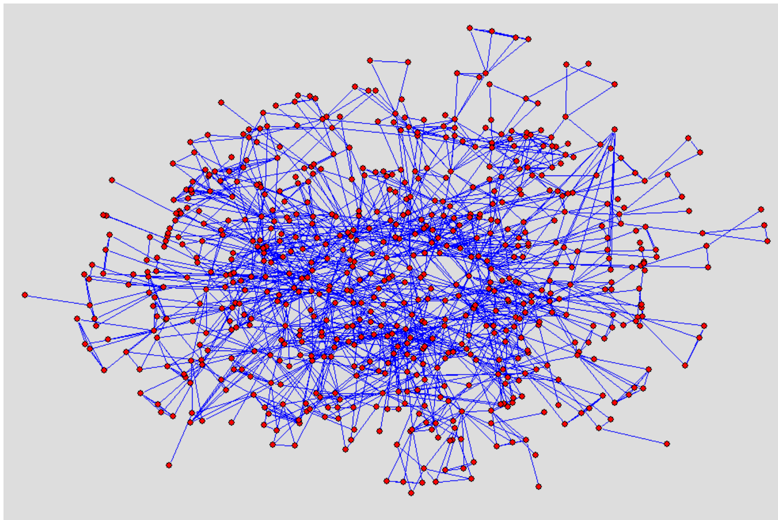
One of our goals in social network analysis

- Find the intrinsic structures:

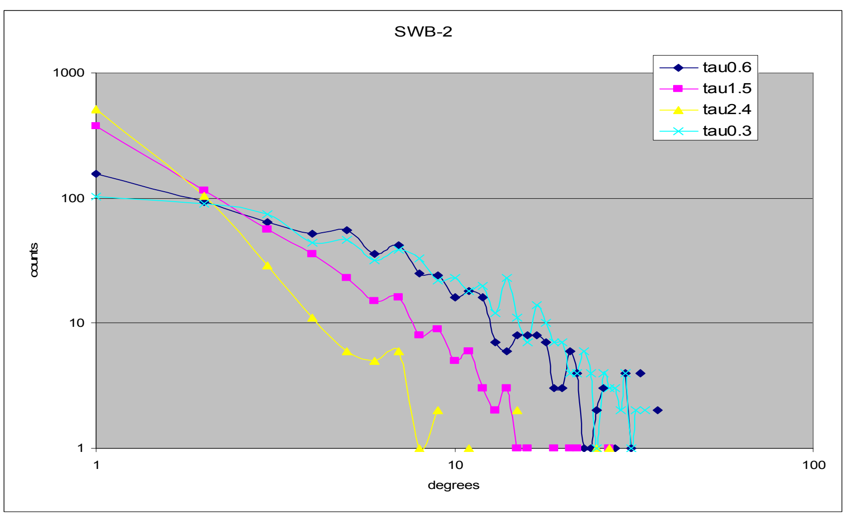


Simulated Intrinsic Bipartite 679 Nodes Social Network

- Nodes 679, cluster coefficient = 0.233



Simulated Social Network based on Bipartite Graphs



Our Insight on the Complex Network Topology

- There exists many intrinsic multipartite properties in the real world.
- Complex network is usually the *outermost* layer of observation.
- To effective synthesize complex networks, it is important to assume the underlying multipartite structure as well as its parameters.
- It may not easy to reconstruct the intrinsic multipartite structure only based on the network topology. Need more information may be needed: e.g., node/user behavior on the effects of different types of data.

Outline

- Complex Network: Characteristics and Examples
- Dynamic Probabilistic Complex Network
- Information Flow in Dynamic Probabilistic Complex Network
- Summary and Conclusion

The Most Difficult Challenge: State-of-the-Arts?

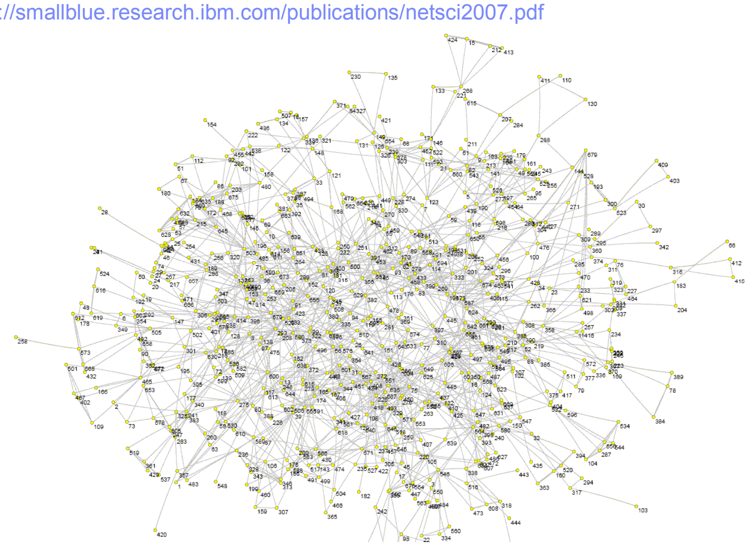
➔ Our Objectives: Find important people, community structures, or information flow in a network, which is *dynamic*, *probabilistic* and *complex*, in order allocate resources in a large-scale mining system.

- Social Networks in sociological and statistic fields: focus on (1) overall network characteristics, (2) dynamic random graphs, (3) binary edges, etc. ➔ Not consider probabilistic nodes/edges or individual nodes/edges.
- Epidemic Networks & Computer Virus Network: focus on (1) overall network characteristics – when will an outbreak occurs, (2) regular / random graphs. ➔ Not focus on individual nodes/edges.
- (Computer) Communication Networks: focus on (1) packet transmission – information is not duplicated, or (2) broadcasting – not considering individual nodes/edges or complex network topology.
- WWW: focus on (1) topology description, (2) binary edges and ranked nodes (e.g., Google PageRank) ➔ Not consider probabilistic edges

What is a Dynamic Probabilistic Complex Network?

- Example: http://smallblue.research.ibm.com

http://smallblue.research.ibm.com/publications/netsci2007.pdf



Modeling a Dynamic Probabilistic Complex Network

- [*Assumption*] A DPCN can be represented by a Dynamic Transition Matrix ***P(t)***, a Dynamic Vertex Status Random Vector ***Q(t)***, and two dependency functions f_M and g_M.

$$\mathbf{P}(t) \triangleq \begin{bmatrix} \mathbf{p_{1,1}}(t) & \mathbf{p_{2,1}}(t) & \cdots & \cdots & \mathbf{p_{N,1}}(t) \\ \mathbf{p_{1,2}}(t) & \mathbf{p_{2,2}}(t) & & & \mathbf{p_{N,2}}(t) \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \mathbf{p_{1,N}}(t) & \mathbf{p_{2,N}}(t) & \cdots & \cdots & \mathbf{p_{N,N}}(t) \end{bmatrix}, \quad \mathbf{Q}(t) \triangleq \begin{bmatrix} \mathbf{q_1}(t) \\ \mathbf{q_2}(t) \\ \vdots \\ \vdots \\ \mathbf{q_N}(t) \end{bmatrix},$$

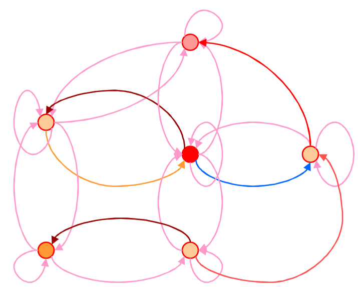
$$\mathbf{P}(t+\delta t) \triangleq f_M(\mathbf{Q}(t), \mathbf{P}(t)),$$

$$\mathbf{Q}(t+\delta t) \triangleq g_M(\mathbf{P}(t+\delta t), \mathbf{Q}(t), \mathbf{P}(t)),$$

$$\mathbf{p_{i,j}}(t) \triangleq \begin{bmatrix} \Pr(y_{i,j}(t) = SE_1) \\ \Pr(y_{i,j}(t) = SE_2) \\ \vdots \\ \Pr(y_{i,j}(t) = SE_{\Omega_E}) \end{bmatrix}, \quad \mathbf{q_i}(t) \triangleq \begin{bmatrix} \Pr(x_i(t) = SV_1) \\ \Pr(x_i(t) = SV_2) \\ \vdots \\ \Pr(x_i(t) = SV_{\Omega_V}) \end{bmatrix},$$

$$\sum_{\omega \in \Omega_E} \Pr(y_{i,j}(t) = SE_\omega) = 1, \quad \sum_{\omega \in \Omega_V} \Pr(x_i(t) = SV_\omega) = 1,$$

$x_i(t)$: the status value of vertex i at time t.
$y_{i,j}(t)$: the status value of edge $i \rightarrow j$ at time t.



Modeling a Dynamic Probabilistic Complex Network – cont'd

- Also the Network Topology should follow the characteristics of complex network:

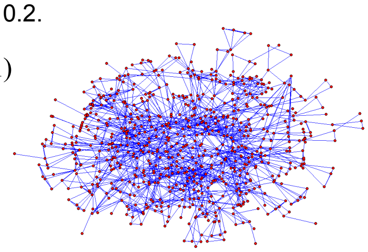
Network topology follows power-law:

$$\Pr(\sum_i \mathrm{u}(p_{i,j}) = l) \sim S \cdot l^{-d} \qquad \mathrm{u}(p_{i,j}) = \begin{cases} 1, & if \exists t, \Pr(y_{i,j}(t) \neq null) > 0 \\ 0, & else \end{cases}$$

d is typically in the range of $2 \sim 2.5$.

and the clustering coefficient C is typically > 0.2.

$$C = \Pr(\mathrm{u}(p_{j,k}) = 1 \mid \mathrm{u}(p_{i,j}) = 1, \mathrm{u}(p_{i,k}) = 1)$$



Modeling a Binary DPCN of binary nodes and edges

- A Binary DPCN can be represented by a Dynamic Transition Matrix *P(t)*, a Dynamic Vertex Status Random Vector *Q(t),* and two dependency functions f_M and g_M.

$$\mathbf{P}(t) \triangleq \begin{bmatrix} p_{1,1}(t) & p_{2,1}(t) & \cdots & \cdots & p_{N,1}(t) \\ p_{1,2}(t) & p_{2,2}(t) & & & p_{N,2}(t) \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ p_{1,N}(t) & p_{2,N}(t) & \cdots & \cdots & p_{N,N}(t) \end{bmatrix}, \quad \mathbf{Q}(t) \triangleq \begin{bmatrix} q_1(t) \\ q_2(t) \\ \vdots \\ \vdots \\ q_N(t) \end{bmatrix},$$

$$\mathbf{P}(t+\delta t) \triangleq f_M(\mathbf{Q}(t), \mathbf{P}(t)),$$

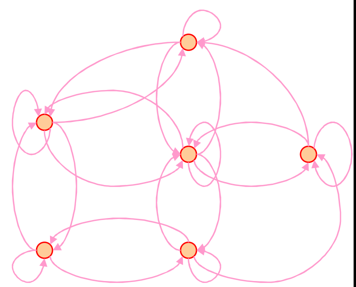
$$\mathbf{Q}(t+\delta t) \triangleq g_M(\mathbf{P}(t+\delta t), \mathbf{Q}(t), \mathbf{P}(t)),$$

$$p_{i,j}(t) \triangleq \Pr(edge_{i->j}(t)=1), \; q_i(t) \triangleq \Pr(x_i(t)=1),$$

$$\Pr(edge_{i->j}(t)=1) + \Pr(edge_{i->j}(t)=0) = 1,$$

$$\Pr(x_i(t)=1) + \Pr(x_i(t)=0) = 1,$$

$x_i(t)$: the status value of vertex i at time t.



Markov Model is a special case of Binary DPCN

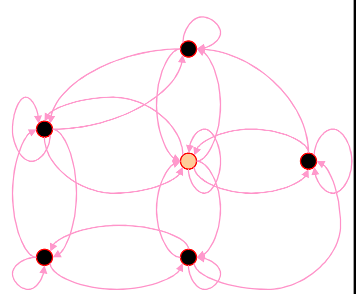
- Markov Model

$$\mathbf{P} \triangleq \begin{bmatrix} p_{1,1} & p_{2,1} & \cdots & \cdots & p_{N,1} \\ p_{1,2} & p_{2,2} & & & p_{N,2} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ p_{1,N} & p_{2,N} & \cdots & \cdots & p_{N,N} \end{bmatrix}, \quad \mathbf{Q}(t) \triangleq \begin{bmatrix} q_1(t) \\ q_2(t) \\ \vdots \\ \vdots \\ q_N(t) \end{bmatrix},$$

$$\mathbf{Q}(t+\delta t) \triangleq g(\mathbf{P}, \mathbf{Q}(t)) = \mathbf{P} \cdot \mathbf{Q}(t)$$

$$\sum_{j=1}^{N} p_{i,j} = 1 \qquad q_i(t) = \Pr(x_i(t)=1)$$

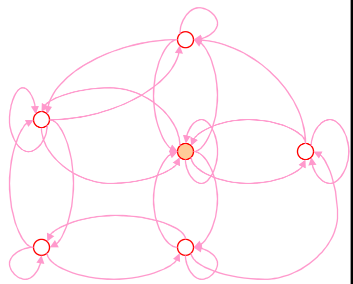
$x_i(t)$: the status value of vertex i at time t.



Many Prior Researches are based on Markov Models

- Random Walks:

$$\lim_{t\to\infty} Q_V(t) = 1-(1-P_E Q_V(0)).*(1-P_E^{(2)} Q_V(0)).*\cdots.*(1-P_E^{(\infty)} Q_V(0))$$

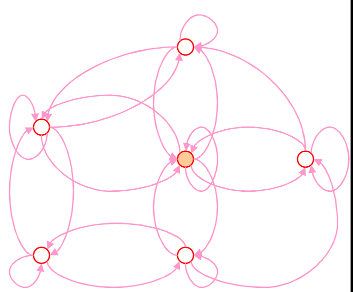


Markov Model is not appropriate to model information flow

- Random Walks assume the existence of a token ➔ unique existence.

- However, information can be duplicated at nodes.

  ➔ New models are needed.



Outline

- Complex Network: Characteristics and Examples
- Dynamic Probabilistic Complex Network
- Information Flow in Dynamic Probabilistic Complex Network
- Summary and Conclusion

Information Flow in Dynamic Probabilistic Complex Network (*Let's call it:* Behavioral Information Flow (BIF) Model)

- [*Assumption*] Edge can be represented by a four-state S-D-A-R (Susceptible-Dormant-Active-Removed) Markov Model. Nodes can be represented by three states S-A-I (Susceptible-Active-Informed) Model.

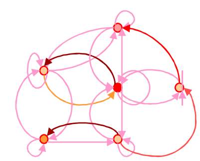
$$\mathbf{P}(t) \triangleq \begin{bmatrix} \mathbf{p_{1,1}}(t) & \mathbf{p_{2,1}}(t) & \cdots & \cdots & \mathbf{p_{N,1}}(t) \\ \mathbf{p_{1,2}}(t) & \mathbf{p_{2,2}}(t) & & & \mathbf{p_{N,2}}(t) \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \mathbf{p_{1,N}}(t) & \mathbf{p_{2,N}}(t) & \cdots & \cdots & \mathbf{p_{N,N}}(t) \end{bmatrix}, \quad \mathbf{Q}(t) \triangleq \begin{bmatrix} \mathbf{q_1}(t) \\ \mathbf{q_2}(t) \\ \vdots \\ \vdots \\ \mathbf{q_N}(t) \end{bmatrix},$$

$$\mathbf{P}(t+\delta t) \triangleq f(\mathbf{M}, \mathbf{Q}(t), \mathbf{P}(t)),$$

$$\mathbf{Q}(t+\delta t) \triangleq g(\mathbf{P}(t+\delta t), \mathbf{Q}(t), \mathbf{P}(t)),$$

$$\mathbf{p_{i,j}}(t) = \begin{bmatrix} \Pr(y_{i,j}(t)=S) \\ \Pr(y_{i,j}(t)=D) \\ \Pr(y_{i,j}(t)=A) \\ \Pr(y_{i,j}(t)=R) \end{bmatrix} \triangleq \begin{bmatrix} \sigma_{i,j} \\ \psi_{i,j} \\ \mu_{i,j} \\ \rho_{i,j} \end{bmatrix}, \quad \mathbf{q_i}(t) = \begin{bmatrix} \Pr(x_i(t)=S) \\ \Pr(x_i(t)=A) \\ \Pr(x_i(t)=I) \end{bmatrix} \triangleq \begin{bmatrix} \lambda_i \\ \eta_i \\ \nu_i \end{bmatrix},$$

$$\sigma_{i,j} + \psi_{i,j} + \mu_{i,j} + \rho_{i,j} = 1 \qquad \lambda_i + \eta_i + \nu_i = 1$$

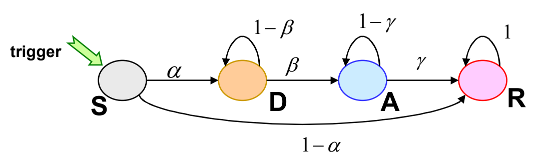


Major Difference between BIF and Prior Modeling Methods in Epidemic Research and Computer Virus Fields

- Model Human Nodes as S-I-R (Susceptible, Infected, and Removed).
- Did not consider individual node's behavior distinctly in network structure/topology ➔ did not consider edge status.
- We propose to model edge status as (autonomous) S-D-A-R Markov Model (Susceptible, Dormant, Active, Removed)
- We propose to model human node behavior as S-A-I (Susceptible, Active, and Informed).

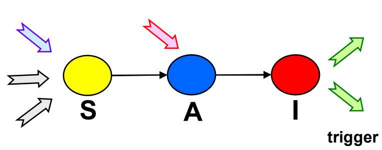
Edges are Markov State Machines, Nodes are not

- State transitions of edges: S-D-A-R model. (Susceptible, Dormant, Active, and Removed) This indicates the time-aspect changes of the state of edges.

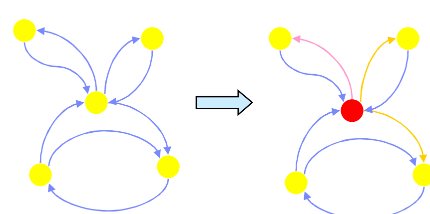


Edge view

- States of nodes: S-A-I model. (Susceptible, Active, and Informed) Trigger occurs when the start node of the edge changes from state S to state I :



Node view



Network view

Edge State Probability and Network Configuration Model

- Nodes and Edges

$$\mathbf{P}(t+\delta t) = f(\mathbf{M}, \mathbf{Q}(t), \mathbf{P}(t)),$$

- Network Configuration Model (which is learned by training). It includes the network topology information, long-term edge probability, and delay parameter).

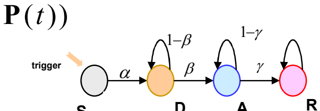
$$\mathbf{M} \triangleq \begin{bmatrix} (\alpha_{1,1},\beta_{1,1},\gamma_{1,1}) & (\alpha_{2,1},\beta_{2,1},\gamma_{2,1}) & \cdots & \cdots & (\alpha_{N,1},\beta_{N,1},\gamma_{N,1}) \\ (\alpha_{1,2},\beta_{1,2},\gamma_{1,2}) & (\alpha_{2,2},\beta_{2,2},\gamma_{2,2}) & & & (\alpha_{N,2},\beta_{N,2},\gamma_{N,2}) \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ (\alpha_{1,N},\beta_{1,N},\gamma_{1,N}) & (\alpha_{2,N},\beta_{2,N},\gamma_{2,N}) & \cdots & \cdots & (\alpha_{N,N},\beta_{N,N},\gamma_{N,N}) \end{bmatrix},$$

  - $\alpha_{i,j} = 0$ ➔ No Edge between i and j
  - Our KDD 2005 paper is a special case that $\alpha_{i,j}$ =1 or 0, and did not model $(\beta_{i,j}, \gamma_{i,j})$

Define Edge State Probability Update Function

Edge State Probability Update function *s.t*

$$\mathbf{P}(t+\delta t) = f(\mathbf{M}, \mathbf{Q}(t), \mathbf{P}(t))$$



- Given three different cases:
  1. On trigger: $x_i(t-\delta t) \neq I, x_i(t) = I$

$$\mathbf{p_{i,j}}(t+\delta t) = \begin{bmatrix} \sigma'_{i,j} \\ \psi'_{i,j} \\ \mu'_{i,j} \\ \rho'_{i,j} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \alpha_{i,j} & 1-\beta_{i,j} & 0 & 0 \\ 0 & \beta_{i,j} & 1-\gamma_{i,j} & 0 \\ 1-\alpha_{i,j} & 0 & \gamma_{i,j} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{i,j} \\ \psi_{i,j} \\ \mu_{i,j} \\ \rho_{i,j} \end{bmatrix} \triangleq \mathbf{F} \cdot \mathbf{p_{i,j}}(t),$$

  2. No trigger – node not informed yet: $x_i(t-\delta t) \neq I, x_i(t) \neq I$

$$\mathbf{p_{i,j}}(t+\delta t) = \mathbf{p_{i,j}}(t),$$

  3. No trigger – node has been informed: $x_i(t-\delta t) = I, x_i(t) = I$

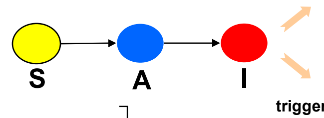
$$\mathbf{p_{i,j}}(t+\delta t) = \mathbf{F} \cdot \mathbf{p_{i,j}}(t),$$

- Therefore, consider the probabilities of node states, then we get $f(.)$:

$$\mathbf{p_{i,j}}(t+\delta t) = \nu_i \cdot \mathbf{F} \cdot \mathbf{p_{i,j}}(t) + (1-\nu_i) \cdot \mathbf{p_{i,j}}(t)$$

Nodes: State Transitions Determined by Incoming Edges

$$\mathbf{Q}(t+\delta t) = g(\mathbf{P}(t+\delta t), \mathbf{Q}(t), \mathbf{P}(t)),$$

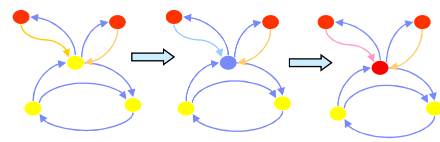


- Node State Probability Update Function $g(.)$:

$$\mathbf{q_i}(t+\delta t) = \begin{bmatrix} \lambda'_i \\ \eta'_i \\ \nu'_i \end{bmatrix} = \begin{bmatrix} \prod_{n\in\Omega_{V,i}} (1-\mu'_{n,i}) & 0 & 0 \\ 1-\prod_{n\in\Omega_{V,i}} (1-\mu'_{n,i}) & \prod_{n\in\Omega_{V,i}} (1-\gamma_{n,i})\mu_{n,i} & 0 \\ 0 & 1-\prod_{n\in\Omega_{V,i}} (1-\gamma_{n,i})\mu_{n,i} & 1 \end{bmatrix} \begin{bmatrix} \lambda_i \\ \eta_i \\ \nu_i \end{bmatrix} \triangleq \mathbf{Q}\cdot\mathbf{q_i}(t),$$

$$\Pr(\exists n \in \{1\ldots N\}, y_{n,i}(t+\delta t) = R, y_{n,i}(t) = A)$$
$$= 1 - \prod_{n\in\Omega_{V,i}} (1-\gamma_{n,i})\mu_{n,i}$$

and $\Omega_{V,i}$ is the set of all source nodes of the incoming edges of Node i: $\Omega_{V,i} = \{n \mid \forall n \in \{1\ldots N\}, \alpha_{n,i} > 0\}$



Network view

Two special considerations for information propagation behavior

- No Reverse Propagation:
  - Add an update criteria to f(.):

$$if\ (y_{i,j}(t+\delta t) = R, y_{i,j}(t) = A)) \Rightarrow y_{j,i}(t+\delta t) = R$$

  - This constraint does not affect $\mathbf{Q}(t)$.
  - It makes the probabilities change to: $\tilde{\psi}'_{j,i} = \tilde{\mu}'_{j,i} = 0$

    and $\tilde{\rho}'_{j,i} = \rho'_{j,i} + \psi'_{j,i} + \mu'_{j,i}$

- No Simultaneously Communication from one person (e.g., phone calls):
  - Add a constraint criteria to f(.):

$$if\ (y_{i,j}(t) = A)) \Rightarrow \forall m \in \Omega_{U,i} \neq j, y_{i,m}(t) \neq A$$

  - And, also: $\Omega_{U,i} = \{m \mid \forall m \in \{1\ldots N\}, \alpha_{i,m} > 0\}$

    where $\Omega_{U,i}$ is the set of all end nodes of the outgoing edges of Node i:
  - The probabilities should be:

$$if\ (y_{k,i}(t) = A)) \Rightarrow \forall n \in \Omega_{V,i} \neq k, y_{n,i}(t) \neq A$$

$$\mathbf{p}_{i,m}(t+\delta t) = \mathbf{p}_{i,m}(t) \quad \text{and} \quad \mathbf{p}_{n,i}(t+\delta t) = \mathbf{p}_{n,i}(t)$$

An Application of Information Flow Prediction – find important people

- Who are the most likely people to talk about this information at a specific time given the current observation?

$$(m,n) = \underset{m,n \in \{1 \ldots N\}}{\arg \max}(\mu_{m,n}(t+\tau)) \text{ given } \mathbf{Q}(t) \text{ or } (\mathbf{P}(t), \mathbf{Q}(t))$$

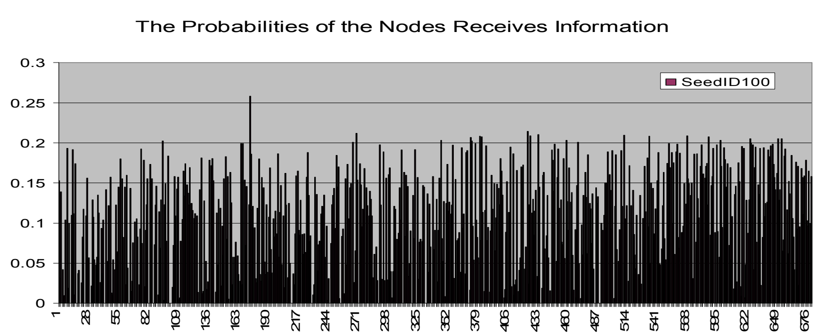
- For a given concrete observation, the values in the given priors $\mathbf{P}(t), \mathbf{Q}(t)$ are either 0 or 1.

- For speaker recognition results, the priors can be confidence values between 0 ~ 1.

Predicting behavioral information flow – Algorithm I

- Monte Carlo Method: Simulate each DPCN information flow for 1000 times.

- It takes 12 seconds to use MC simulation to predict the process. (For a given model and test 679 nodes, it takes a PC 130 mins for calculate the probabilities if the information flow starts from different 679 seeds).



Outline

- Complex Network
- Dynamic Probabilistic Complex Network
- Information Flow in Dynamic Probabilistic Complex Network
  – Who should we monitor? Where to put the sniffers?
  – Training, Effect of Noises, and Dynamic Model Updates in Dynamic Probabilistic Complex Network
- Demo
- Next Steps
  – Communities in Dynamic Probabilistic Complex Network

Training the Network Configuration

- Train the Network Configuration Model **M** using the observation data in Time $0 - T$:

$$\alpha_{i,j} = \frac{L}{K}$$

$$\beta_{i,j} = \frac{1}{E[w]}$$

$$\gamma_{i,j} = \frac{1}{E[d]}$$

where K is the number of times that node x_i becomes active during time period $0\text{-}T$. L is the count of the number of times that edge $y_{i,j}$ becomes active.

In the Markov Model of SDAR, the duration of the information staying in the D state is a Poisson distribution with mean value = $1/\beta_{i,j}$. We can then estimate the parameter $\beta_{i,j}$ based on the mean waiting time $E[w]$ of training data. Similarly, we can get $\gamma_{i,j}$ based on the mean active duration $E[d]$ of training data.

Impact of Classification Error on BIF Model

- Consider two types of errors:
  - –Speaker Recognition Error
    - E.g. DIG scenario ➔ nodes may have miss and false alarm. This types of error would cause edge error.
  - –Topic Detection Error
    - E.g. classification of email content ➔ nodes are correct. Edges may have miss and false alarm.
  - –Combination of both errors
    - E.g., if we are doing both topic detection and speaker recognition in DIG, then the above two types of error will be combined.

An Application of Information Flow Prediction – enhance speaker recognition accuracy

- The probability that a given pair talks about this information at a specific time given the current observation?

$$\Pr(y_{i,j}(t+\tau) = A) = \mu_{i,j}(t+\tau) \quad \text{given } \mathbf{Q}(t) \text{ or } (\mathbf{P}(t), \mathbf{Q}(t))$$

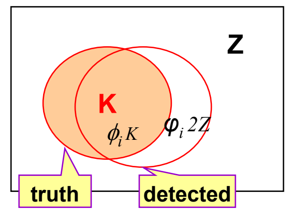
- These probabilities can serve as prior confidence for speaker recognition.

Noise Factor I – Impact of Classification Error from Speaker Recognition

- Assume the classification precision rate on the speaker (node) i is ϕ_i, and the false alarm rate on the speaker i is φ_i.
- Then the expected number of times that the node is counted is:

$$\tilde{L} = \phi_i \phi_j L + \varphi_i \varphi_j Z$$

- And the link is counted is: $\tilde{K} = \phi_i K + \varphi_i \cdot 2Z$



- Therefore,

$$\tilde{\alpha}_{i,j} = \frac{\tilde{L}}{\tilde{K}} = \frac{\phi_i \phi_j L + \varphi_i \varphi_j Z}{\phi_i K + \varphi_i \cdot 2Z}$$

- If we assume a universal precision and false alarm rate at all speakers, then:

$$\tilde{\alpha}_{i,j} = \frac{\tilde{L}}{\tilde{K}} = \frac{\phi^2 L + \varphi^2 Z}{\phi_i K + \varphi_i \cdot 2Z}$$

Assume the average waiting time of links and the average transmission duration of links are the same regardless of the links observed, then:

$$\tilde{\beta}_{i,j} = \beta_{i,j} \quad \text{and} \quad \tilde{\gamma}_{i,j} = \gamma_{i,j}$$

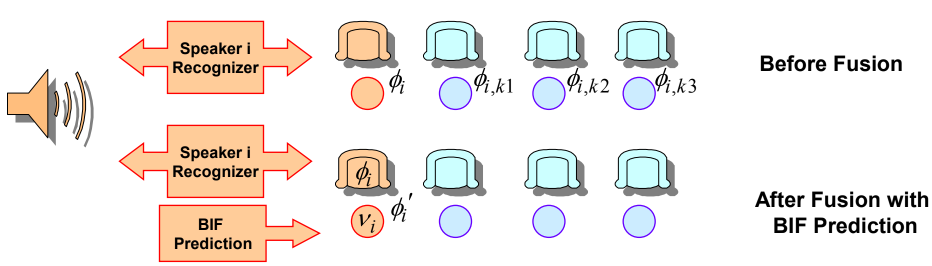
- If we assume the false alarm rate is small and can be neglected when the number of nodes is large, then $\tilde{\alpha}_{i,j} \approx \phi \cdot \alpha_{i,j}$

Speaker Recognition Accuracy can be Improved by Fusion of Original Speaker Recognition and Predicted Node Probability

- We can use this fusion method to combine both speaker recognition result and the estimated node probability:

$$\phi_i' = \frac{\phi_i \cdot \nu_i}{\phi_i \cdot \nu_i + \sum_k \phi_{i,k} \cdot \nu_k}$$

which is guaranteed to be increasing when $\nu_i > \forall \nu_k$



Dynamic Updates on Speaker Recognition Result and the BIF Model

- Assume a special case that a speaker is usually mistakenly classified as the other speaker. E.g., given a true Speaker i speaking, her voice is sometimes classified as Speaker k. (but not the reverse direction)
- Let (n) represents the *n-th* dynamic update of the model. Each time the model is updated based on Slide 'Impact of Classification Error on Model'
- Based on the previous slide, we shall get:

$$\phi_{(n)} = \frac{\phi \cdot v_{i(n)}}{\phi \cdot v_{i(n)} + (1-\phi) \cdot v_{k(n)}} \triangleq \frac{\phi}{\phi + (1-\phi) \cdot \kappa_{(n)}} \quad \text{where} \quad \kappa_{(n)} \triangleq \frac{v_{k(n)}}{v_{i(n)}}$$

- Based on the previous two slides, we can get: $\kappa_{(n)} = \frac{1-\phi_{(n-1)}}{\phi_{(n-1)}} \cdot \kappa_{(n-1)}$
- This value can be calculated as: $\kappa_{(n)} = \left[\frac{1-\phi}{\phi}\right]^n \cdot \kappa_{(0)}^{2n}$

  which quickly converges to zero. ➔ $\lim_{n\to\infty} \phi_{(n)} = 1$

Dynamic Updates on Speaker Recognition Result and the Model – cont'd

- Assume another special case that a speaker is usually mistakenly classified as the other speaker. E.g., given a true Speaker i speaking, her voice is sometimes classified as Speaker k. And, also, Speaker k's voice can be confused as Speaker i.
- Following similar steps as in the previous slide, we shall get:

$$\phi_{(n)} = \frac{\phi \cdot v_{i(n)}}{\phi \cdot v_{i(n)} + (1-\phi) \cdot v_{k(n)}} \triangleq \frac{\phi}{\phi + (1-\phi) \cdot \kappa} \quad \text{where} \quad \kappa_{(n)} = \kappa$$

- If we assume the confusion error is not uniformly the same, i.e., asymmetric error between speakers, then:

$$\phi_{i,(n)} = \frac{\phi_i \cdot v_{i(n)}}{\phi_i \cdot v_{i(n)} + (1-\phi_i) \cdot v_{k(n)}} \triangleq \frac{\phi_i}{\phi_i + (1-\phi_i) \cdot \kappa_{(n)}}$$

$\kappa_{(n)} \triangleq \frac{v_{k(n)}}{v_{i(n)}}$ which depends on the network topology and does not have a closed-form solution.

Noise Factor II – Impact of Classification Error from Topic Classification

- Assume the classification precision rate on the edge i→j is $\theta_{i,j,}$ and the false alarm rate on the edge is $\omega_{i,j}$.
- Then the expected number of times that the edge is counted is:

$$\hat{K}_i = \theta_{i,j} K + \omega_{i,j} Z$$

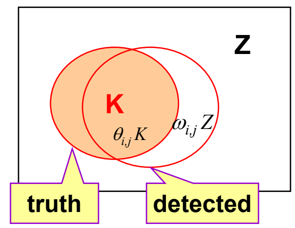
- And the link is counted is:

$$\hat{L} = \theta_{i,j} L + \omega_{i,j} Z$$

- Therefore, $\hat{\alpha}_{i,j} = \frac{\hat{L}}{\hat{K}} = \frac{\theta_{i,j} L + \omega_{i,j} Z}{\theta_{i,j} K + \omega_{i,j} Z}$

- If the false alarm rate can be neglected:

$$\hat{\alpha}_{i,j} = \frac{\theta_{i,j} L}{\theta_{i,j} K} = \alpha_{i,j}$$



Noise Factor II – Impact of Classification Error from Topic Classification – cont'd

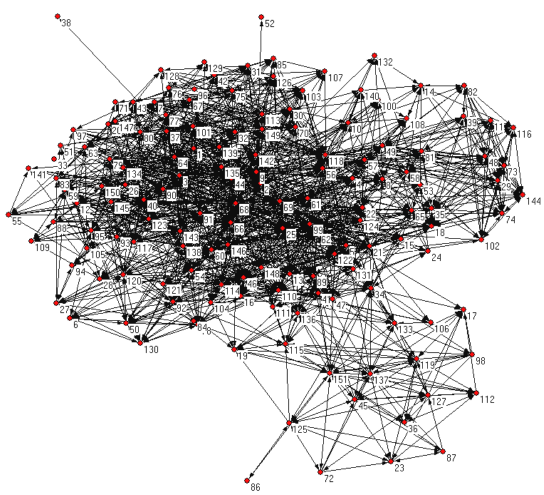
- If the false alarm rate can be neglected ➔ The probability that a classification error occurs at an conversation record is equal to the probability that the nodes are detected.
- Therefore, since the propagation coefficient α's are the same, there will be no effects on the information flow prediction.
- If we consider both the topic classification error and the speaker recognition error together, the information flow prediction will be the same as the case when there is only speaker recognition error.

Multiple Topics

- Each topic is one information flow model. Multiple Topics can be considered as a combination of these models.

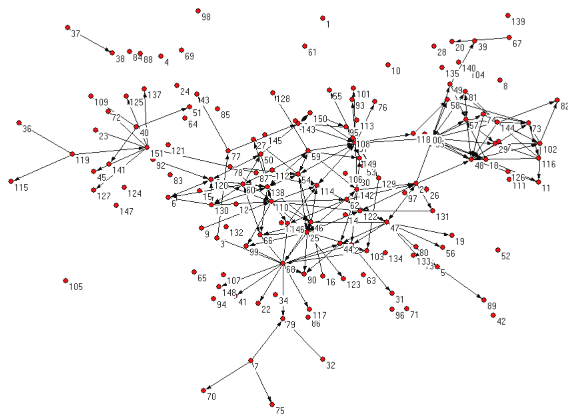
Example: Social Network of Enron Managers

- If we try to build social networks based on communications regardless expertise or topics, it is difficult.



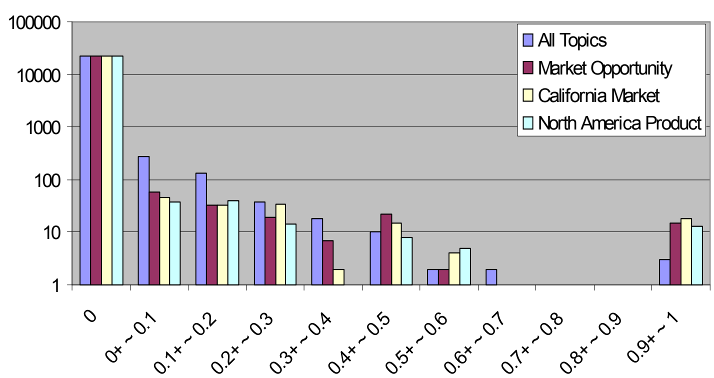
We can first find the experts and then see how this community works (II)

- Rosalee Fleming played an important role at “Market Opportunities.” She received info from Actor 119 (Mike Carson) and Actor 23 (James Steffes – VP of Gov. Affairs of Enron.)
- Actor 68 (Rod Hayslett -- CFO) is also a major information spreader.



We can estimate the parameters in the DPCN model

- Example: a histogram of the alpha values by applying the DPCN model on Enron Dataset.



Questions?