TATA CONSULTANCY SERVICES
Experience certainty.

TATA

TCS Research Scholar Program

Topic – Modeling Information Dissemination in Delay Tolerant Network using Bipartite Network.

Speaker –Sudipta Saha
Indian Institute of Technology, Kharagpur

4th May 2012

HOTEL TAJ PALACE, NEW DELHI

Sudipta Saha

PhD Student

Under the guidance of
Prof Niloy Ganguly

In collaboration with
Prof Animesh Mukherjee

Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur-721302

Financial support by
Tata Consultancy Services

Overview

- **DTN**
  - Mobile handheld devices
  - Wireless, short-range inefficient
  - Intermittent/opportunistic
- **Information Dissemination**
  - store-carry-forward
    - Buffers / Throwbox /Relay Point installed in social places

- **Buffer time is a crucial issue**
  - Overhead
    - The amount of redundant message copies
    - Consumption of battery power
    - Memory consumption
  - **Coverage – No of social places**

Therefore, it is essential to decide the optimal buffer time to achieve a certain coverage for a specified cost

Overview

❑ Analysis of coverage in information dissemination in DTN is difficult using traditional mathematical tools

❑ We show that -

❖The ***coverage*** achieved in the information dissemination process for a given **buffer time** in **DTN**

❖The size of the largest component in the one-mode projection for a certain **varying threshold** in the corresponding **Bipartite Network**

Match with each other

❑ **Existing theory of bipartite network can be exploited to analyze coverage of Information dissemination in DTN**

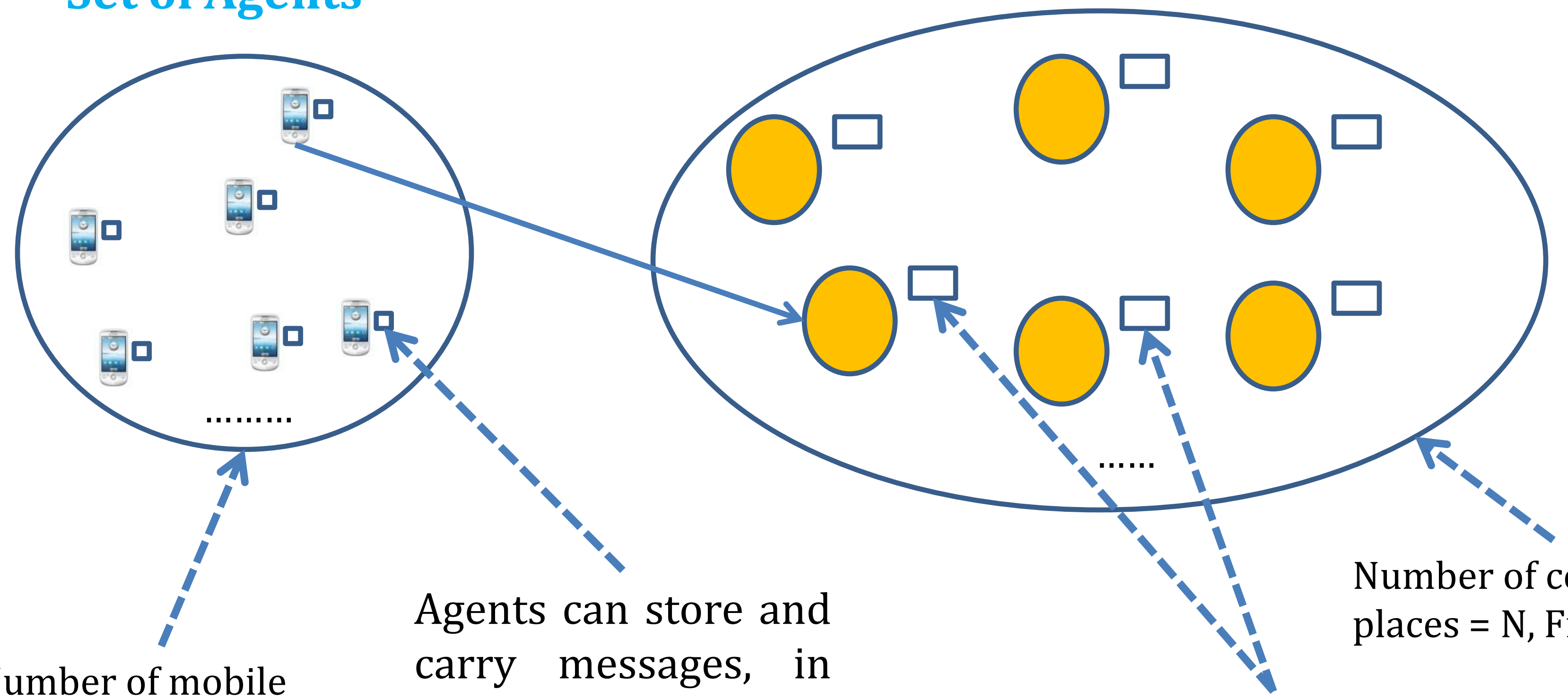
Simplified View of DTN

Each agent visits on average μ places sequentially and preferentially

Set of Agents

Set of common Places

Markets, malls, pubs, stations are examples of common place



Number of mobile agents = t, Growing

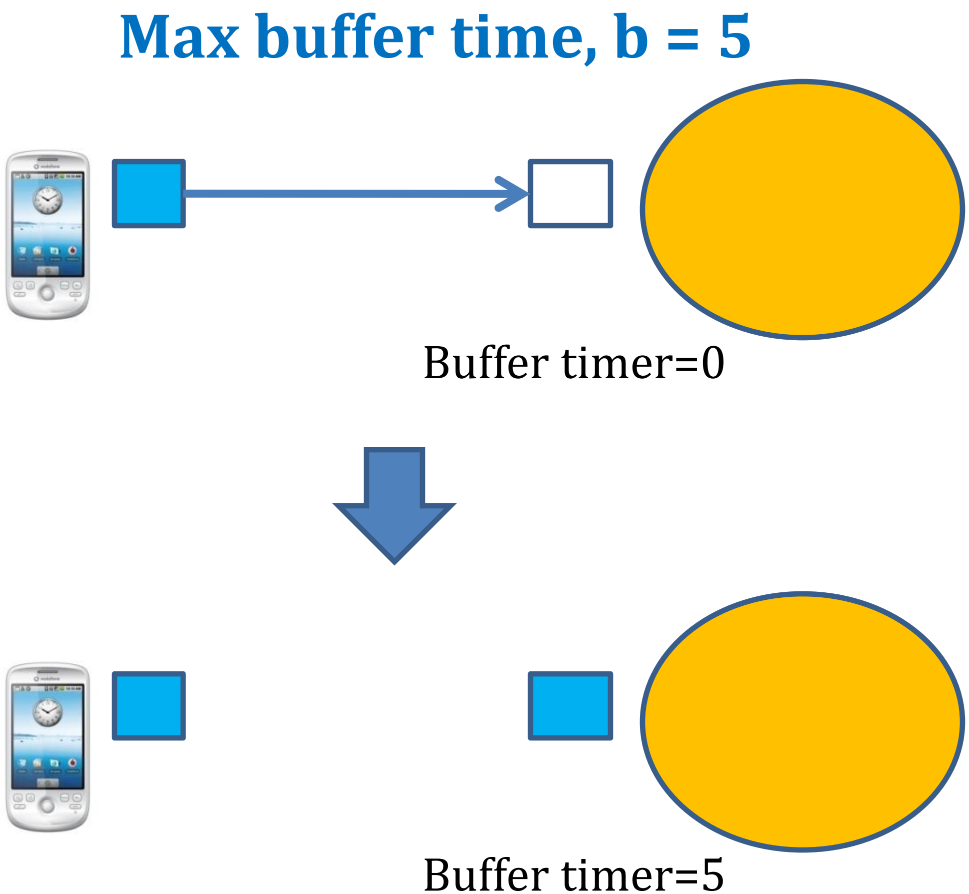
Agents can store and carry messages, in their local device buffer

Each place has a **message buffer** and maintains a **buffer timer**

Number of common places = N, Fixed

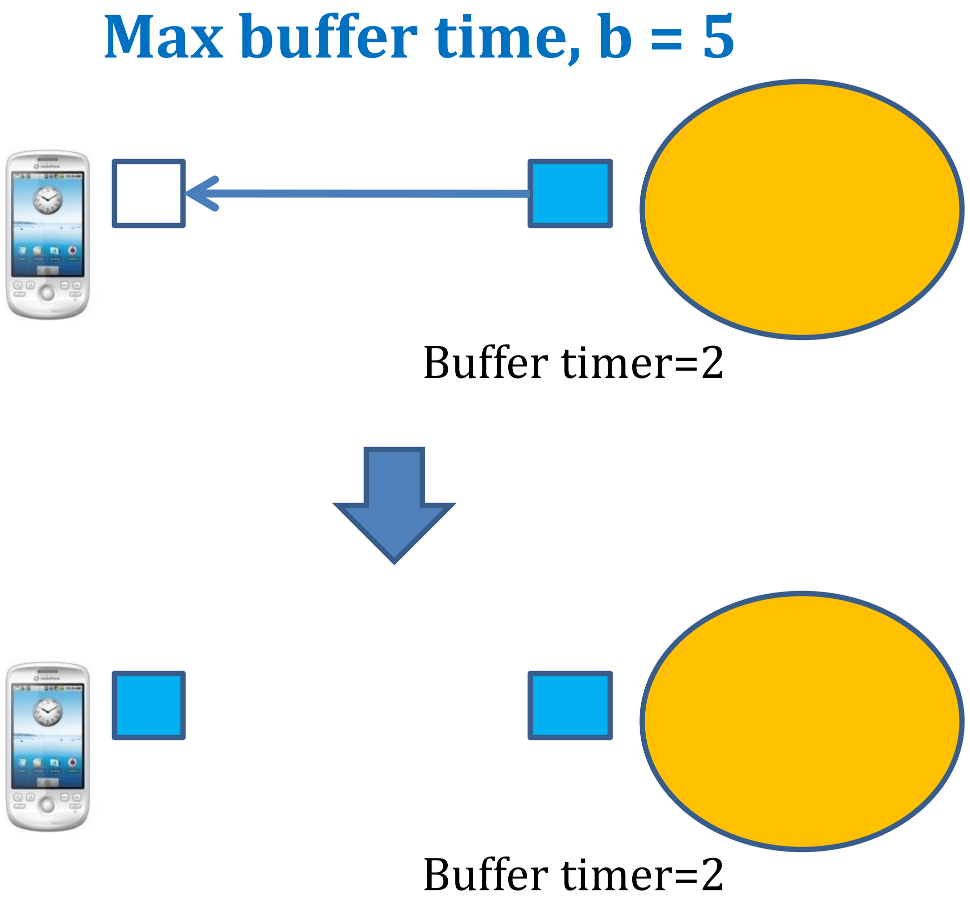
Information Dissemination in DTN

- Case 1
  - ❖ Message transfer from agent to place
  - ❖ Buffer timer in place gets reset to global/max buffer time



Information Dissemination in DTN

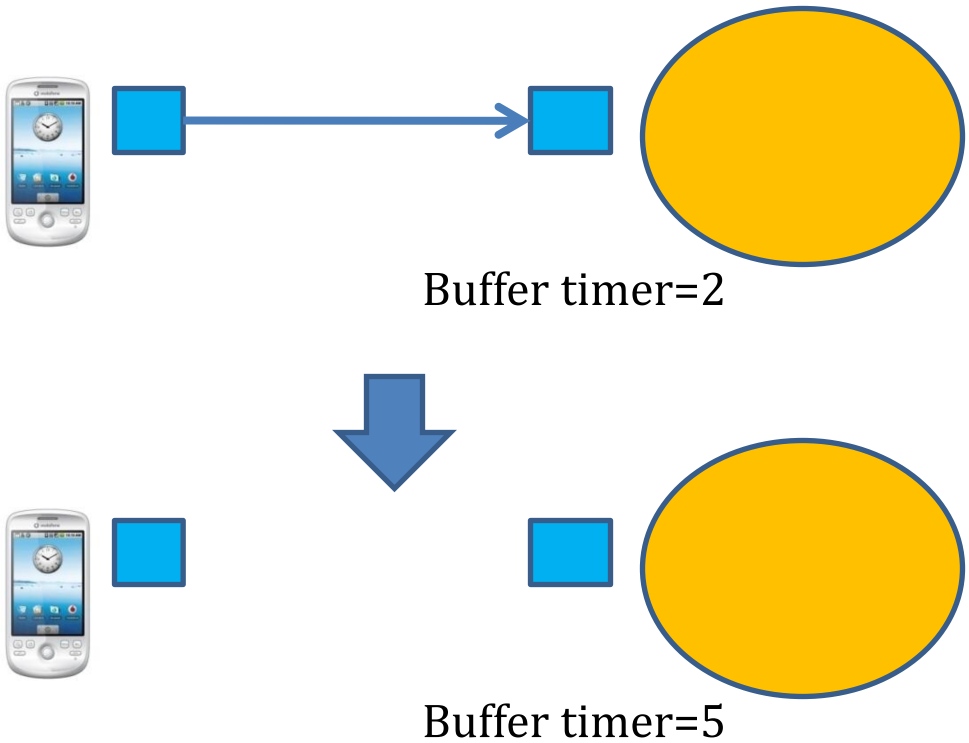
- Case 2
  - Message transfer from place to agent
  - No change in buffer timer



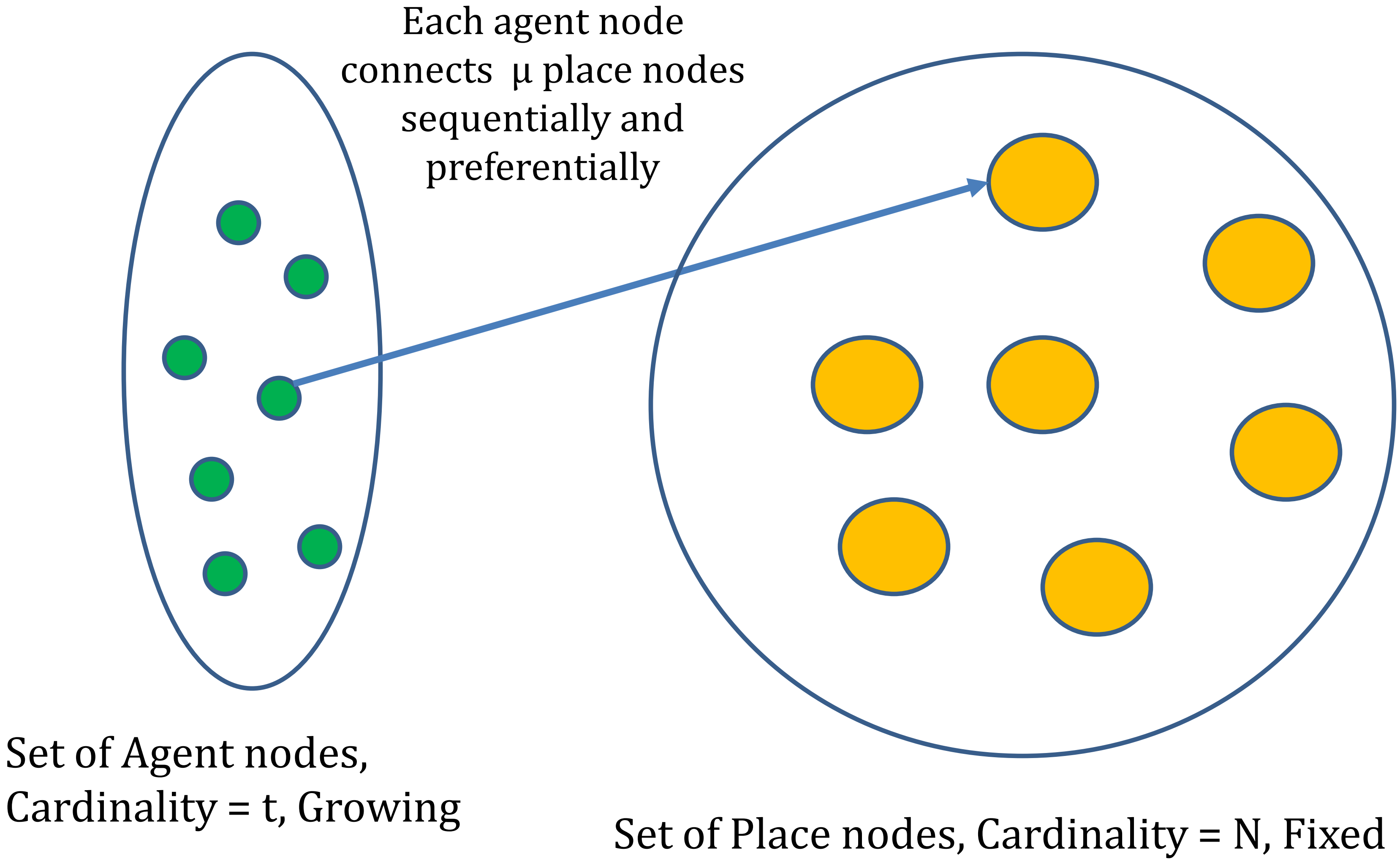
Information Dissemination in DTN

- ❑ Case 3
  - ❖ Agent and place both contain the message
  - ❖ Place is given advantage
    - ○ Buffer timer of place is set to max buffer time, b
- ❑ Case 4 : A trivial case
  - ❖ No message in place or agent
  - ❖ Nothing happens

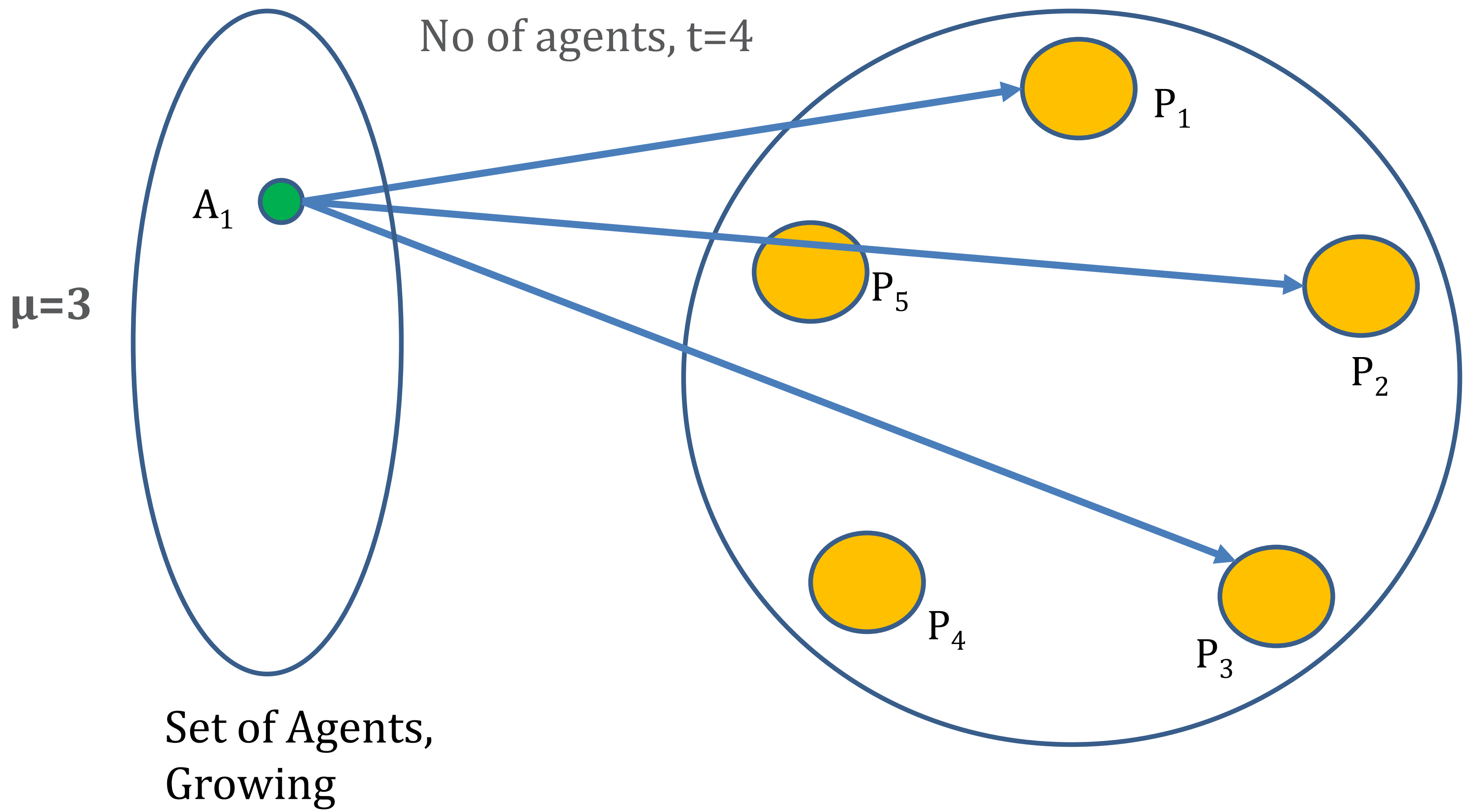
Max buffer time, b = 5



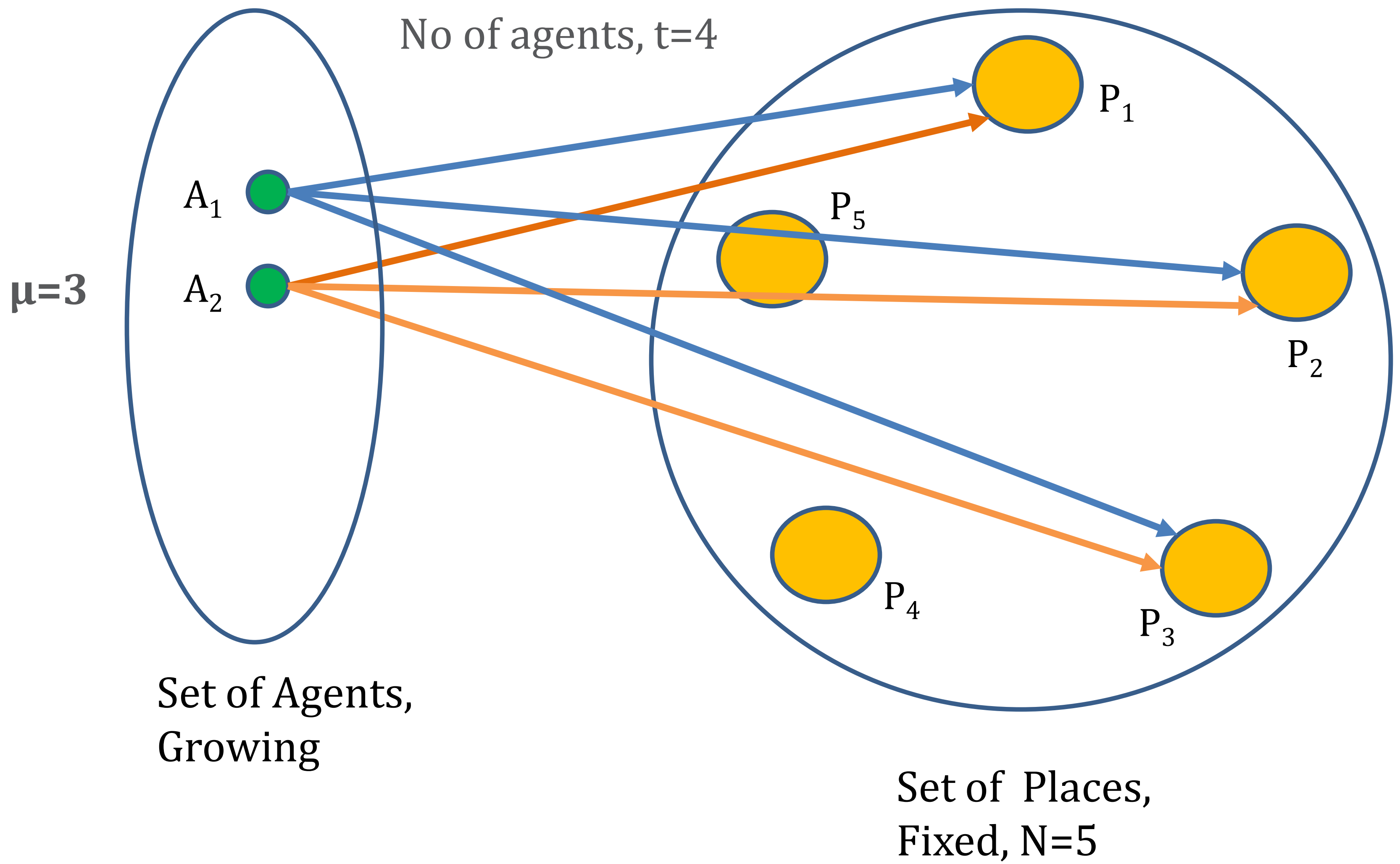
Modeling by Bipartite Network



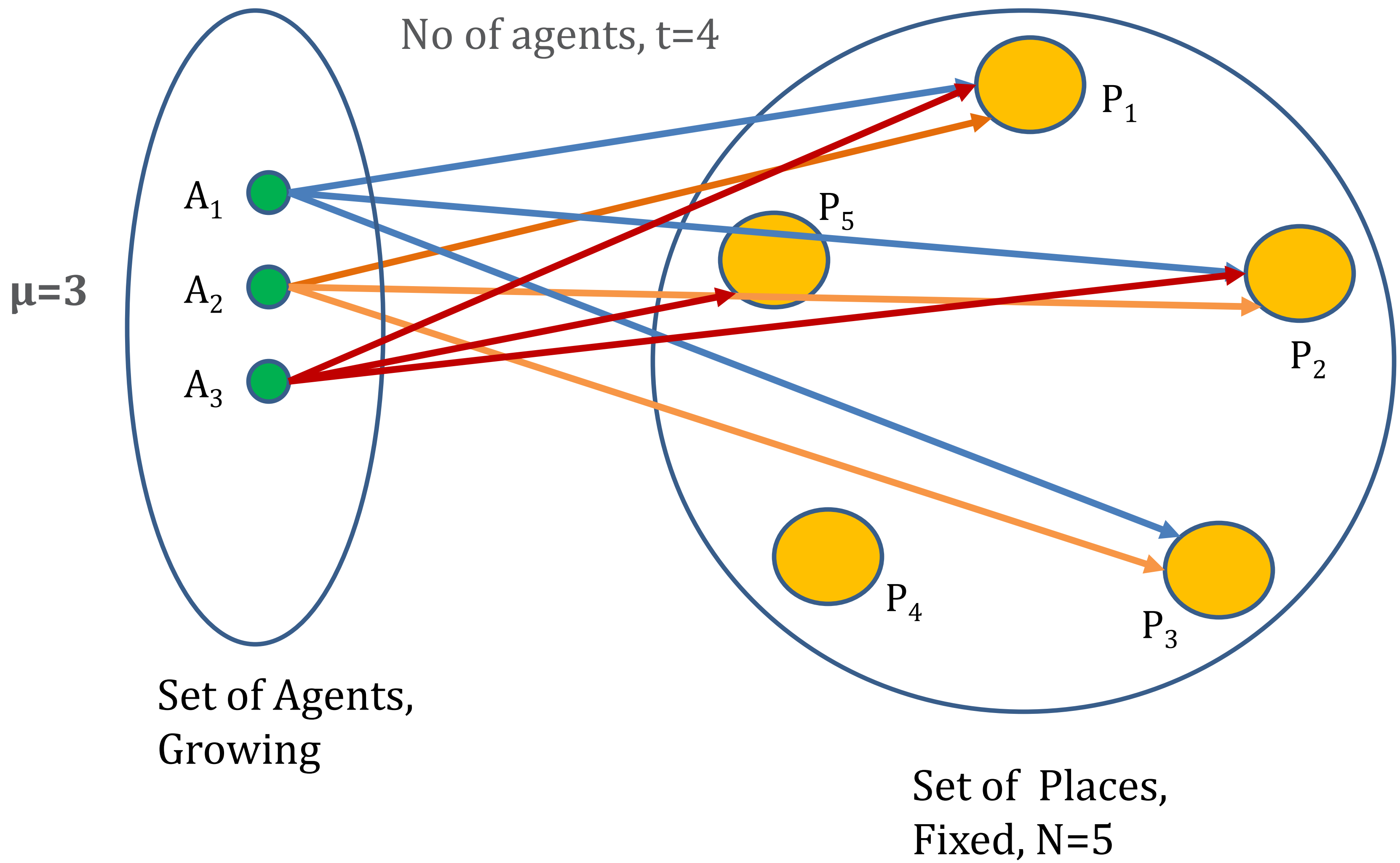
Modeling by Bipartite Network



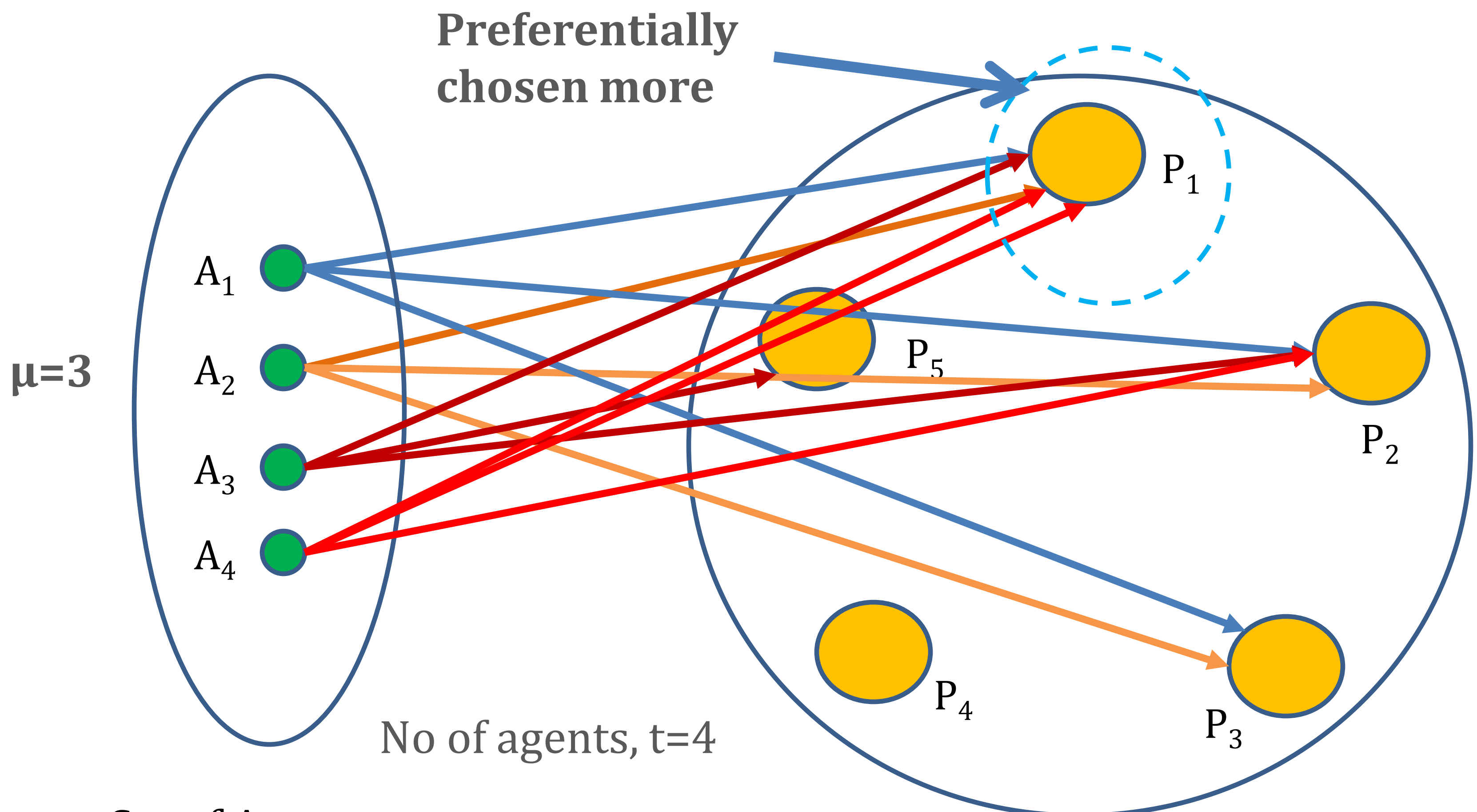
Modeling by Bipartite Network



Modeling by Bipartite Network



Modeling by Bipartite Network

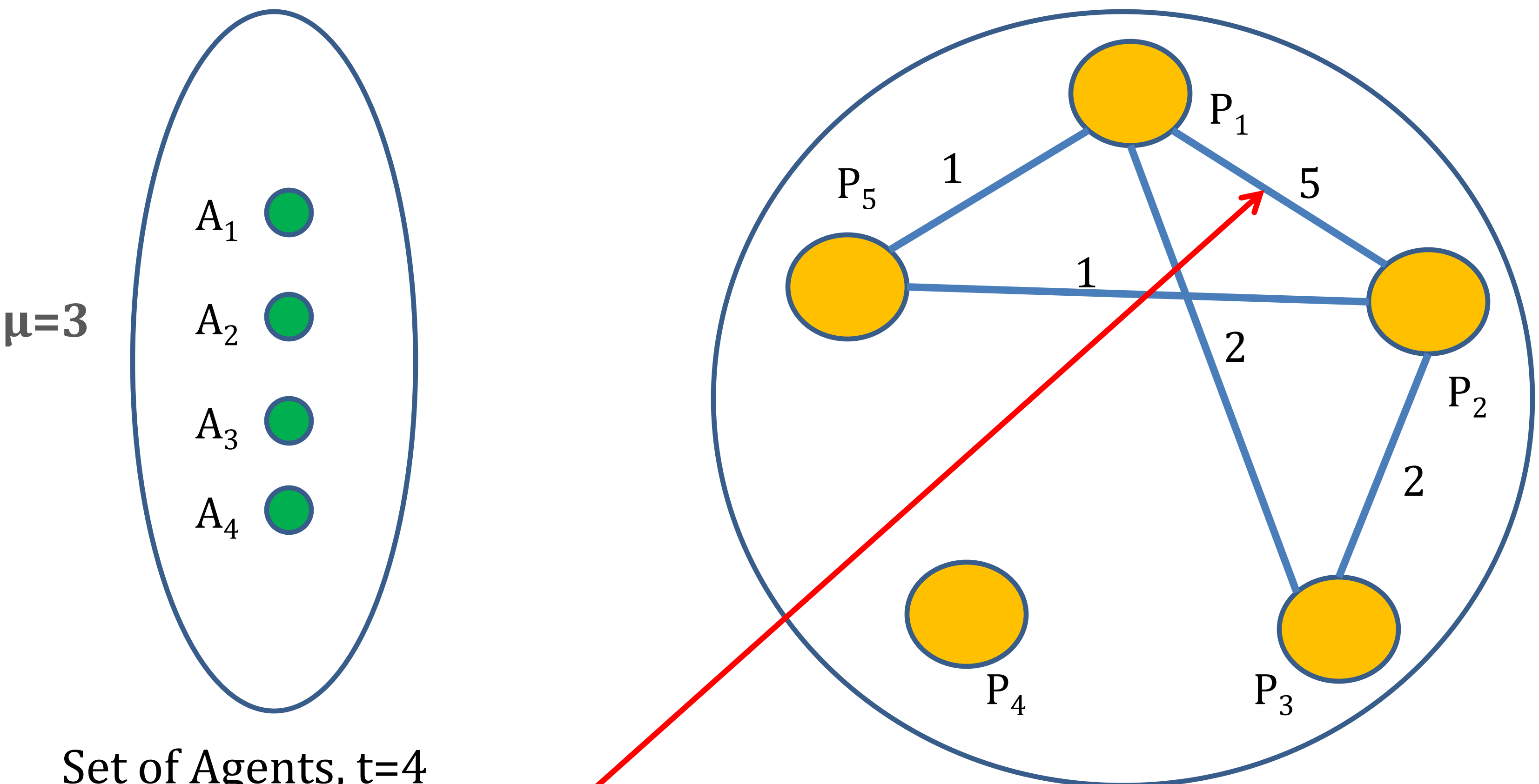


Set of Agents

Set of Places,
Fixed, N=5

Modeling by Bipartite Network

Formation of the one-mode projection



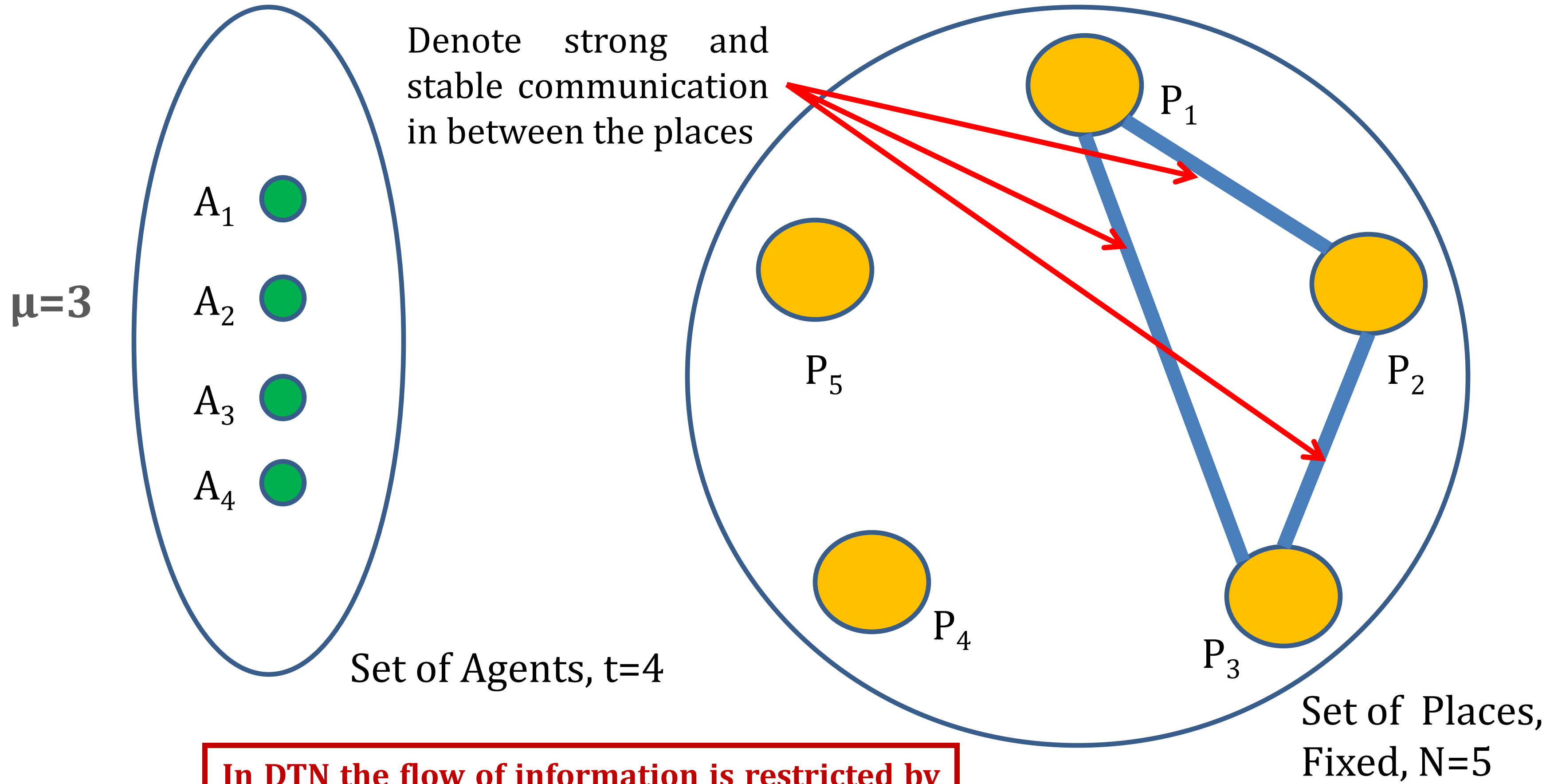
Set of Agents, t=4

Set of Places, N=5

The weight of an edge denotes the number of common visits between two places via same agent

Modeling by Bipartite Network

Thresholded one-mode projection for threshold value = 2



In DTN the flow of information is restricted by buffer time – the rational for thresholding in bipartite network

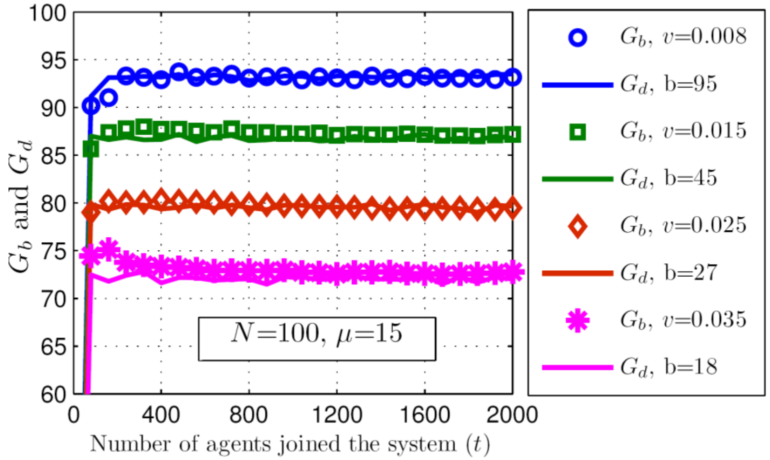
DTN – Bipartite Network : Parameter Mapping

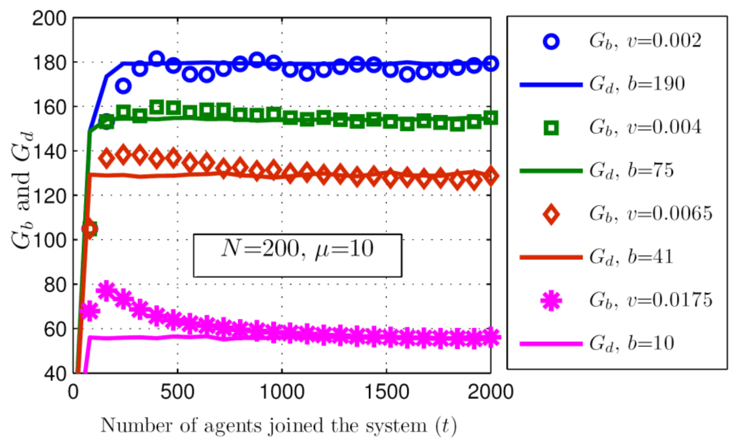
Types	DTN	Bipartite Network	Remarks
Parameters	Agents (t)	Agent partition (t)	Growing
	Places (N)	Place partition (N)	Fixed and finite
	Number of places an agent visits (μ)	Number of connections an agent creates with different places (μ)	Constant (can be taken from some specified distribution also)
	Buffer time (b)	Threshold varying with t (v)	Limitation in buffer time imposes restriction in the achieved coverage
Observable	Coverage, i.e., the number of places where the message could reach under the dissemination process (denoted by G_d)	Size of the largest connected component in the thresholded one-mode projection (denoted by G_b)	These quantities should match

Relationship between Threshold and Buffer Time

- **Buffer time (b) and varying threshold (v) :**
  - Effectiveness of common visits depends on the buffer time (b)
    - Minimum common visits
  - To bring the notion of temporal stability, threshold weight c is calculated as
    - **c = v x t, where v is called a 'time varying threshold' ; more user – more information – probability of one information passing decreases**
  - Relationship between v and b is expressed as follows (A, α and C are constants)

$$v = Ab^{-\alpha} + C$$



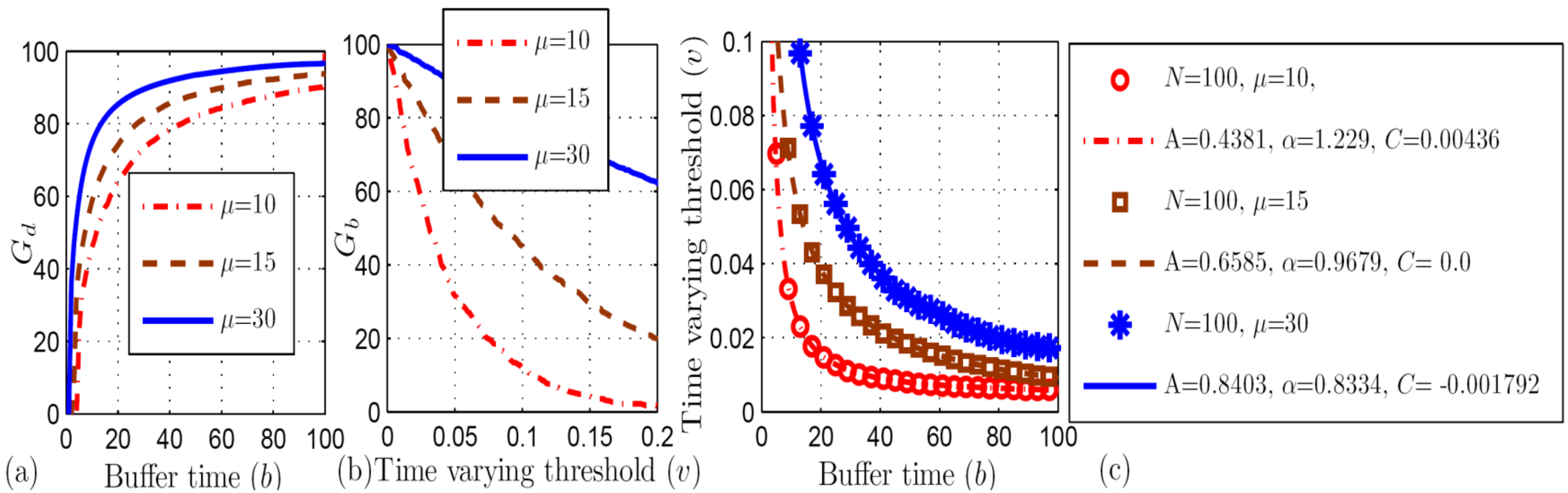


TATA CONSULTANCY SERVICES
Experience certainty.

Relationship between Threshold and Buffer Time

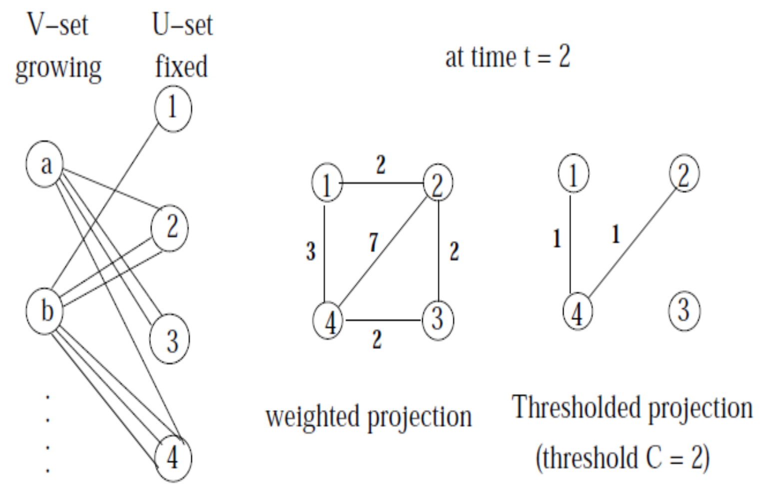
- **Buffer time (b) and varying threshold (v) :**
  - Coverage increases with buffer time (b)
  - Size of the largest component decreases with varying threshold (v)
  - Through extensive simulation and curve fitting we find the following relationship

$$v = Ab^{-\alpha} + C$$



Theory of Bipartite Network

- Set U: fixed number of nodes (set of place nodes), set V : grows with time (set of agents)
- Formal definition of one-mode projection:
  - **Projection G* of bipartite network G = {U,V} onto set U is a uni-partite network containing nodes in U. Nodes x, y ε U are linked in G* if x, y have a common neighbor in V**
  - Un-weighted or weighted projection
- Weighted projection:
  - Connect x, y by as many edges as the number of length-2 paths between x and y in the bipartite network G
- Thresholded projection:
  - Connect x, y by a single edge if there are more than c (which is the **threshold value**) length-2 paths between x and y in bipartite network G



Theory of Bipartite Network

- Closed-form expressions for cumulative degree distribution at large time
  - Fixed partite-set

$$F_t(k) \quad = \quad \left(1 - \frac{k}{\mu t}\right)^{N-1}$$

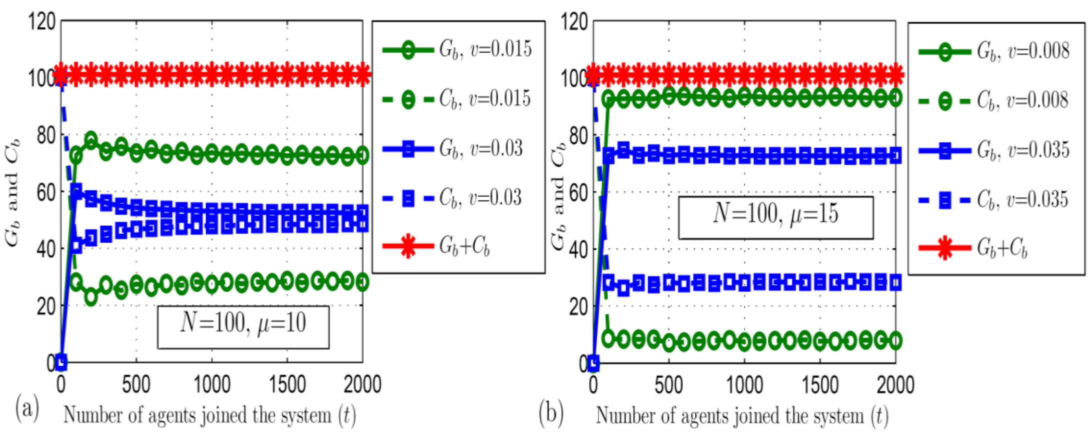
  - Projection onto fixed set at large time

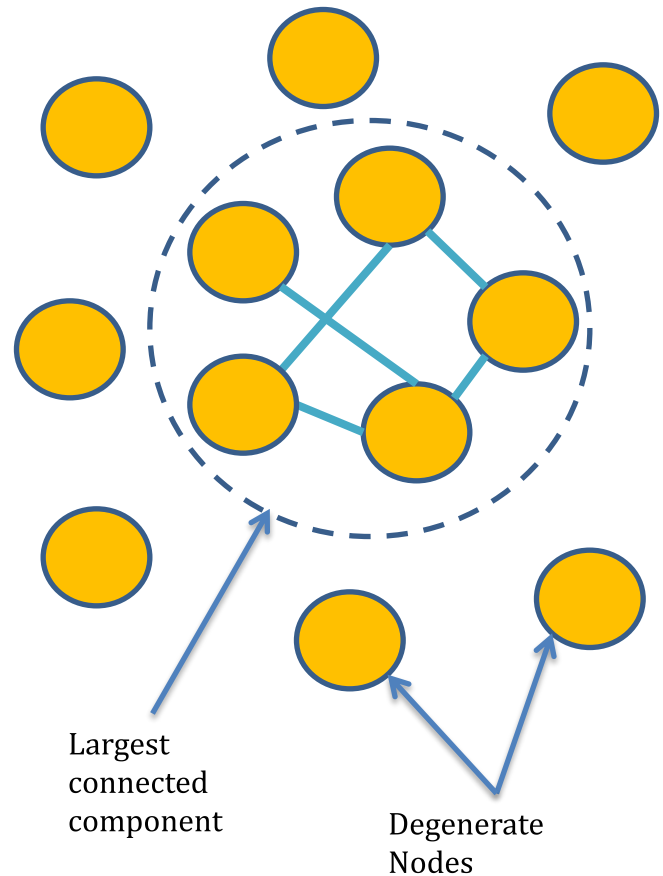
$$F_t(k) \quad = \quad \left(\frac{1+\sqrt{1-4x}}{2}\right)^{N-1} - \left(\frac{1-\sqrt{1-4x}}{2}\right)^{N-1} \qquad \text{where } x = \frac{k}{t(\mu_2^{'}-\mu)}$$

  - Thresholded projection onto the fixed set
    - **We use this result later in coverage estimation**

$$F_t(k) \quad = \quad \left(1 - \frac{c}{(\mu_2^{'} - \mu)x}\right)^{N-1} \qquad x = 1 - \left(\frac{k}{N-1}\right)^{\frac{1}{N-1}}$$

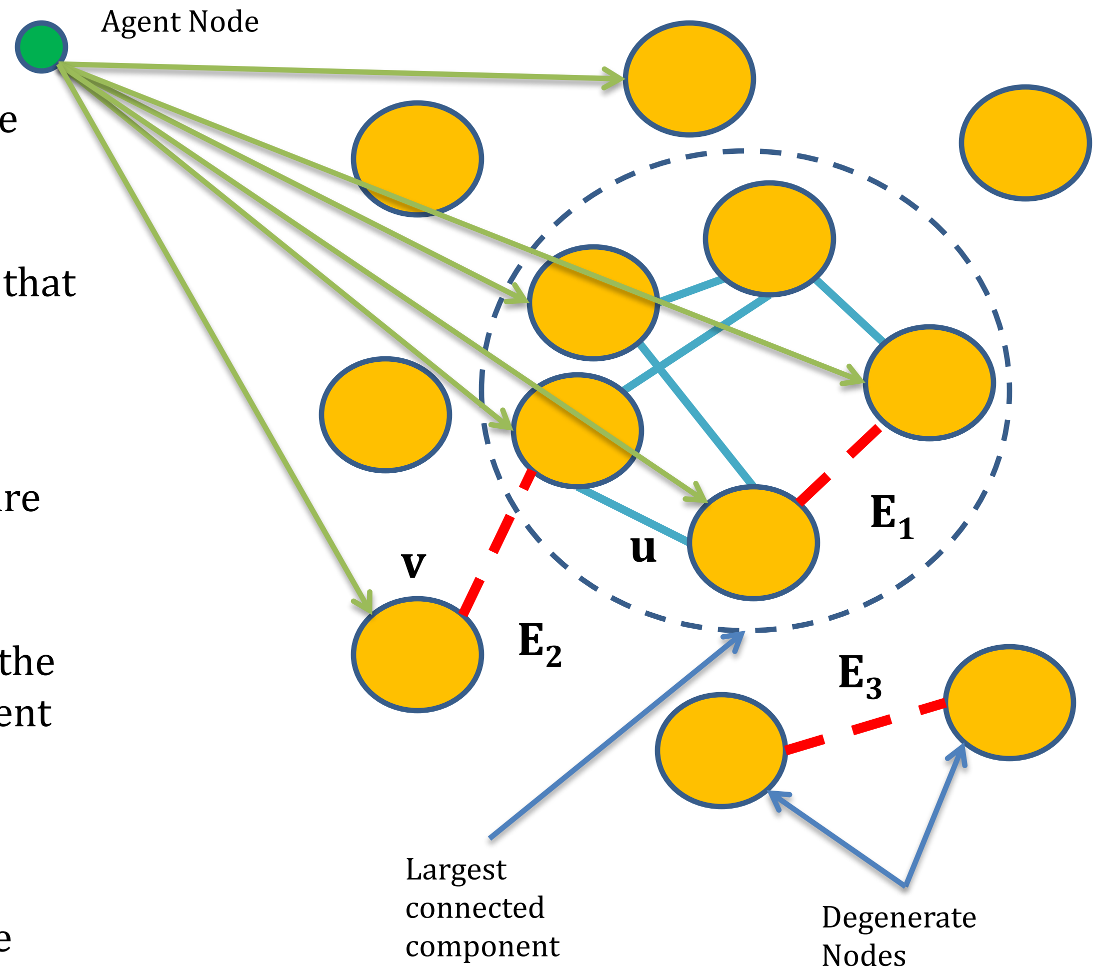
- A special property of the thresholded one-mode projection
  - *After any number of agents have joined, the thresholded one-mode projection of the bipartite network on the place set, consists of a single connected component while the rest of the places that are not part of the largest component are degenerate, i.e., have degree zero. (**Empirically observed for the first time**)*
  - ***Number of Component + size of the largest component = Number of places***





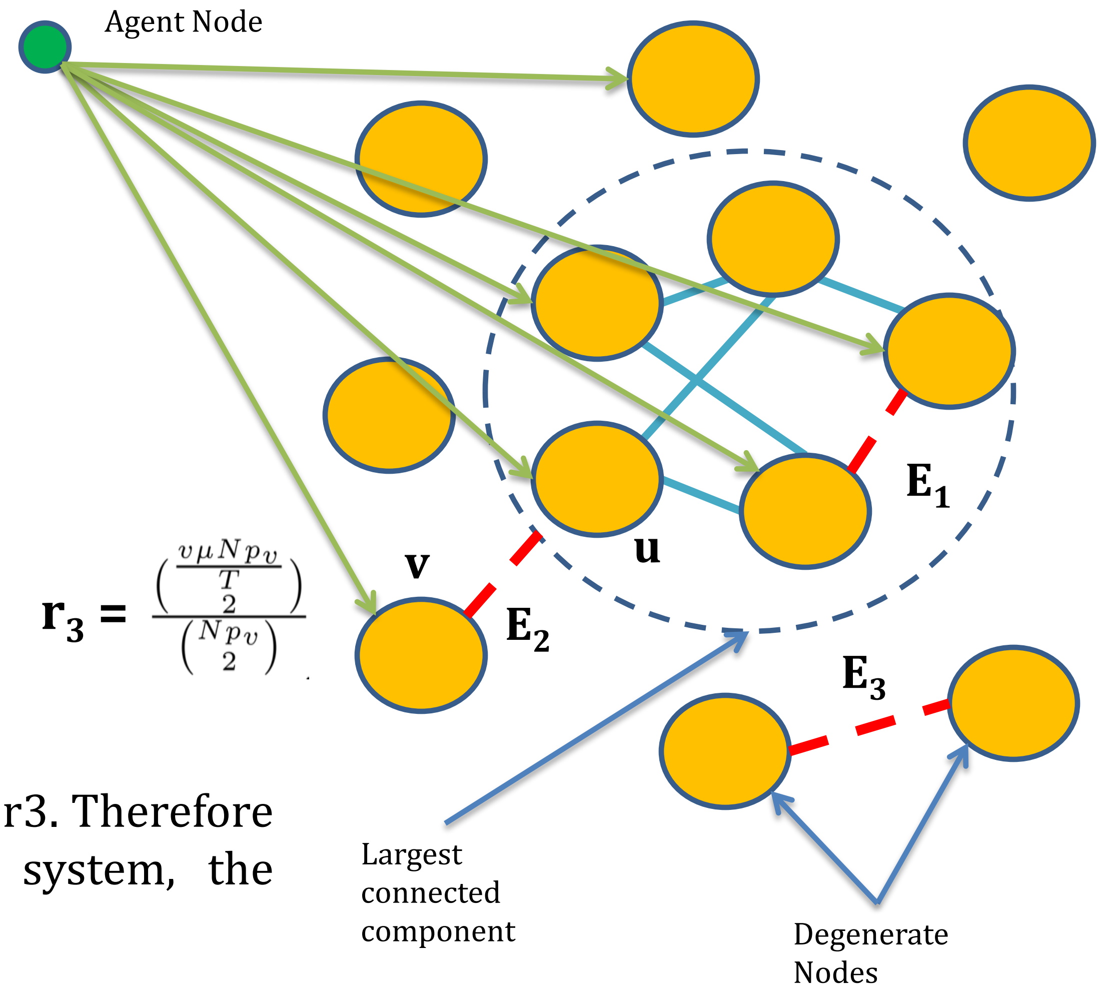
Coverage Estimation

- Let us consider-
  - Two nodes u and v – one inside the giant component and one outside the giant component
  - p_u and p_v are the probabilities that the new agent will create connection with node u and v respectively
  - Let us consider E_1 , E_2 and E_3 are the three sets of edges which connect two nodes
    - Both of which are inside the giant connected component
    - One is inside the giant component and another outside
    - Both of which are outside the giant component



Coverage Estimation

- It can be proved that the rate of increment of the edge weights in these three sets - per agent joining the system – are r1, r2 and r3 respectively – (N is the total number of places)

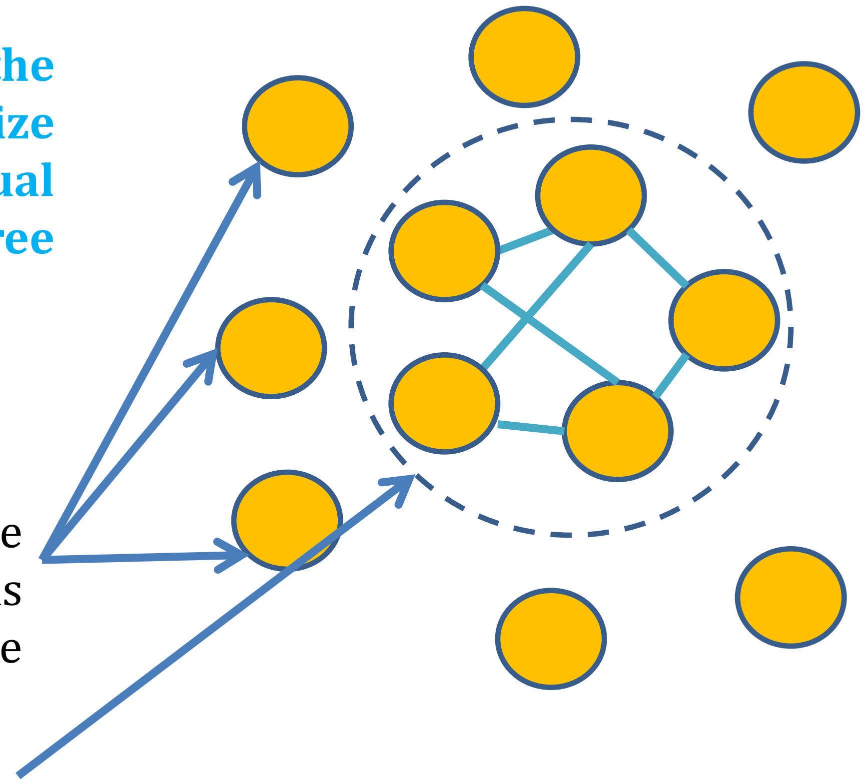


$$\mathbf{r_1} = \frac{\binom{\frac{u\mu N p_u}{T}}{2}}{\binom{N p_u}{2}} \qquad \mathbf{r_2} = \frac{\frac{u\mu N p_u}{T} \times \frac{v\mu N p_v}{T}}{N p_u \times N p_v} \qquad \mathbf{r_3} = \frac{\binom{\frac{v\mu N p_v}{T}}{2}}{\binom{N p_v}{2}}$$

- It can be shown that, r1 > r2 > r3. Therefore after the new agent joins the system, the property is still satisfied

Coverage Estimation

- ❑ **By the virtue of this property, the coverage can be calculated as the size of largest component which is equal to the fraction of nodes with degree greater than zero**



❑ **Number of degenerate nodes**

❖The fraction of these nodes can be found from the degree distribution as the probability that a node has degree zero, i.e., p(0)

❑ **Largest connected component size**

❖The fraction of nodes having degree greater than zero can be calculated from the degree distribution of the thresholded one mode projection by subtracting p(0) as follows : $\mathbf{F_t(0) - p(0) = F_t(1)}$

- We use the formula of degree distribution derived by Ghosh et al. to find this fraction
  - The final form of the size of the largest component is

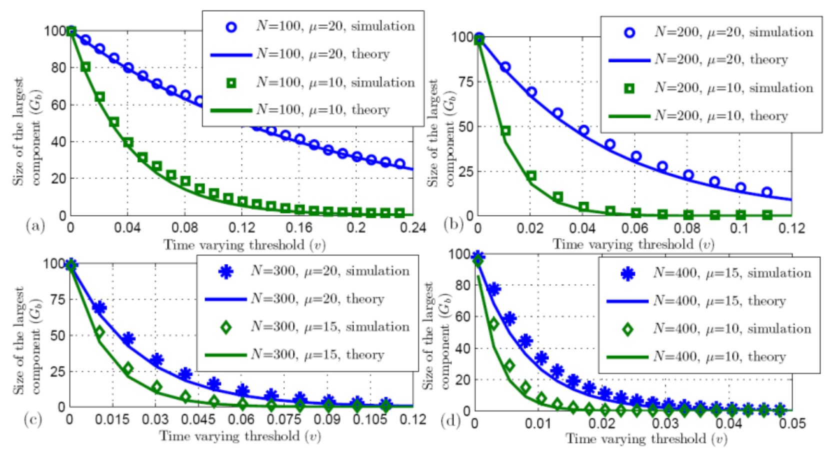
$$G_b = N \times \left[1 - \left(\frac{\sqrt[N-1]{(N-1)}}{\sqrt[N-1]{(N-1)-1}}\right) \times \left(\frac{v}{\mu^2 - \mu}\right)\right]^{N-1}$$

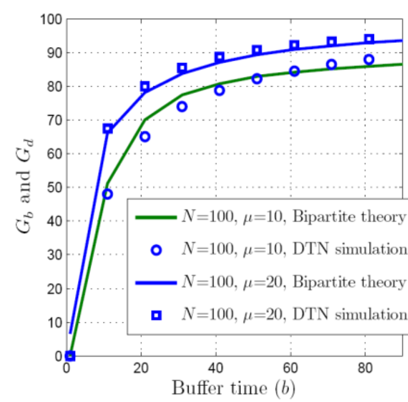
  - Simplified form

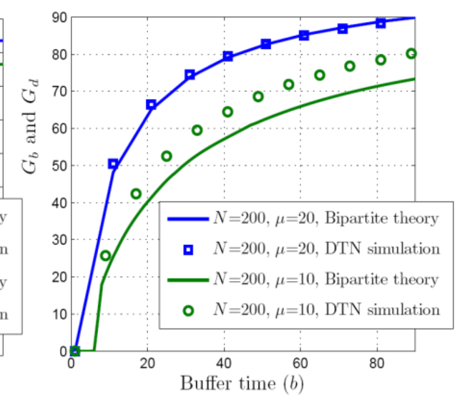
$$G_b = N - \frac{N(N-1)}{\mu(\mu-1)} \times v$$

  - Putting the expression of v we get

$$G_d (= G_b) = N - \frac{N(N-1)}{\mu(\mu-1)} \times (Ab^{-\alpha} + C)$$



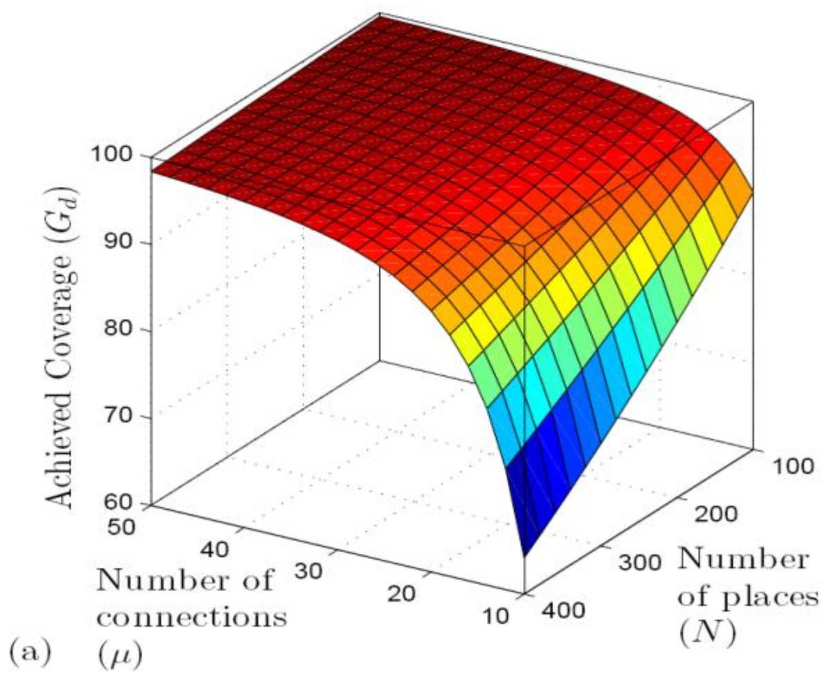




Insights

- The *coverage* is inversely proportional to N^2 and directly to μ^2

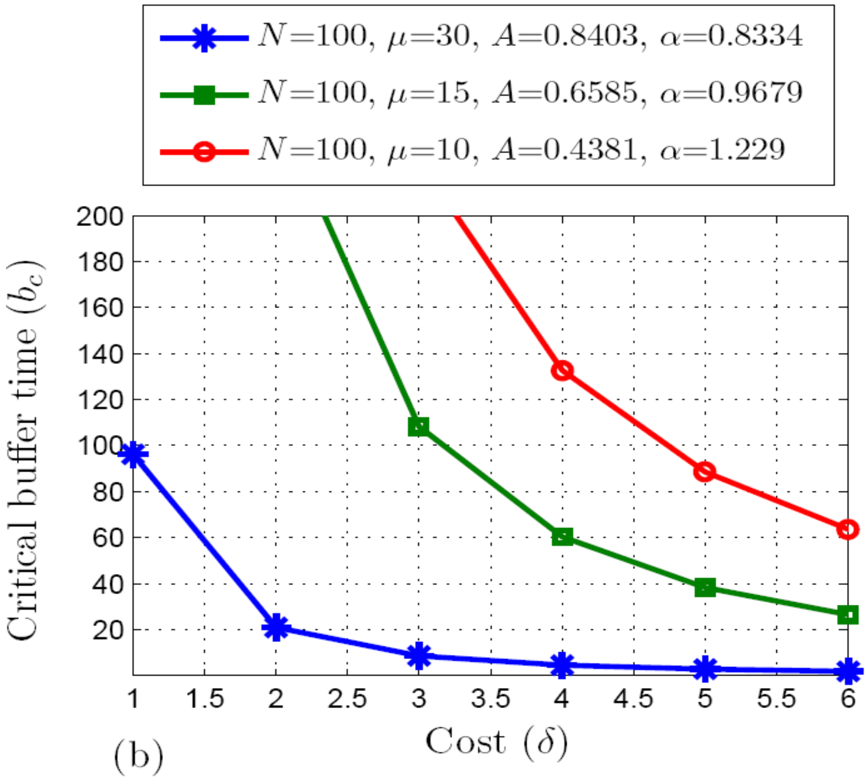
- The *coverage* does not grow unboundedly with the number of agents (*t*) *joining the system*
  - After a certain value of *t*, *the total number of place* nodes covered, gets stabilized and is limited by the buffer time *b*



Insights

- The *coverage* does not grow unboundedly with buffer time (b)
  - ❖After a critical value (say b_c) *coverage is almost stable*
  - ❖*The optimal buffer time can be designed using the following relationship (The rate of increment of coverage with buffer time should be greater than the associated cost increment)*

$$\frac{dG_d}{db} > \delta$$



Benefit/Cost ratio (δ)

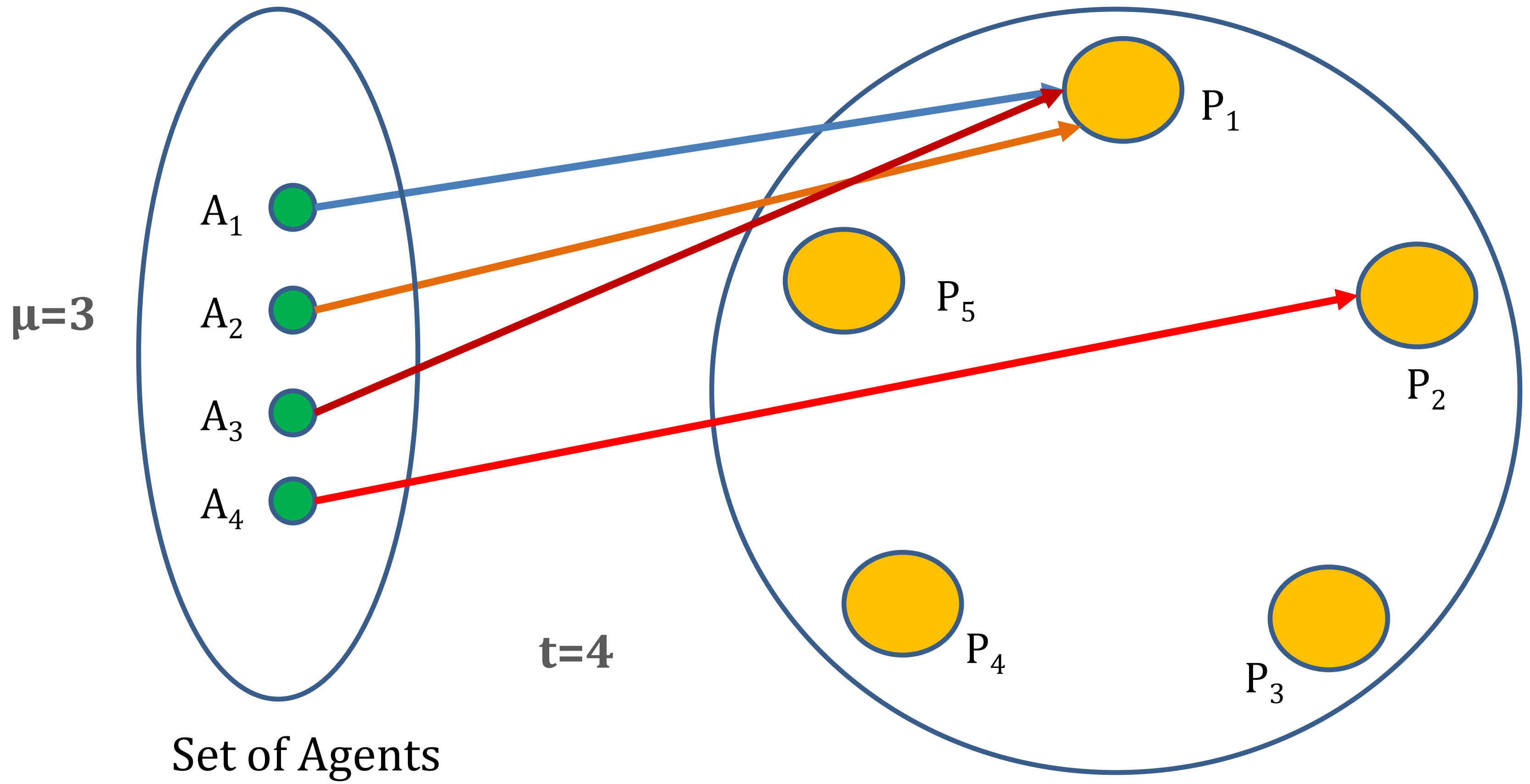
Insights

- We study the effect of the following two more practical aspects in the process of information dissemination in buffer-augmented DTN

  - When there is a bursty nature in the agents' arrival pattern
    - We assume that the life spans of the agents are no longer fully disjoint
    - **Rather, they overlap**; i.e., we now consider that more than one agents can stay in the system in a given time instance

  - Instead of being fully preferential, if there **is some randomness** in the place selection by the agents (explained later)

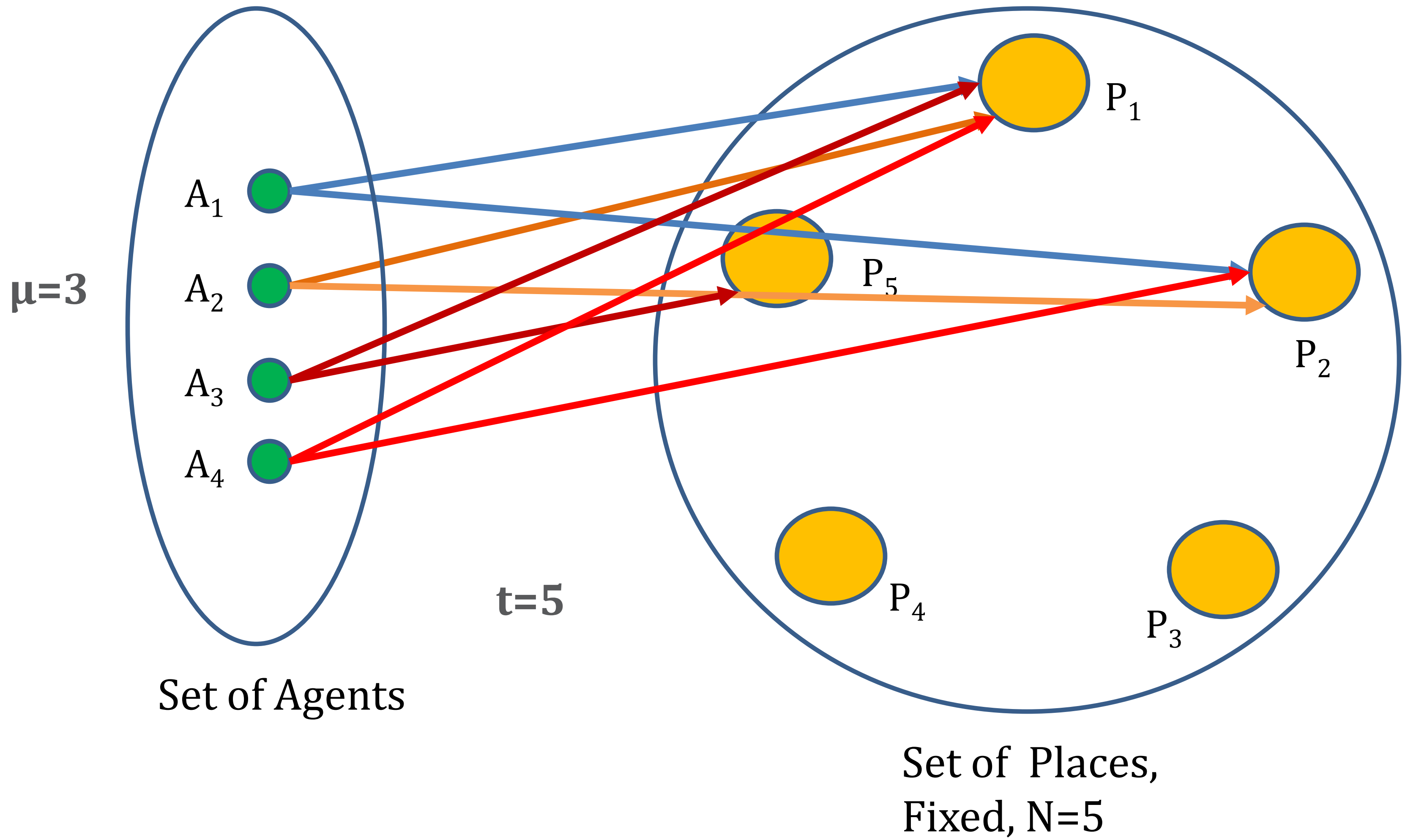
Overlapping Life Span



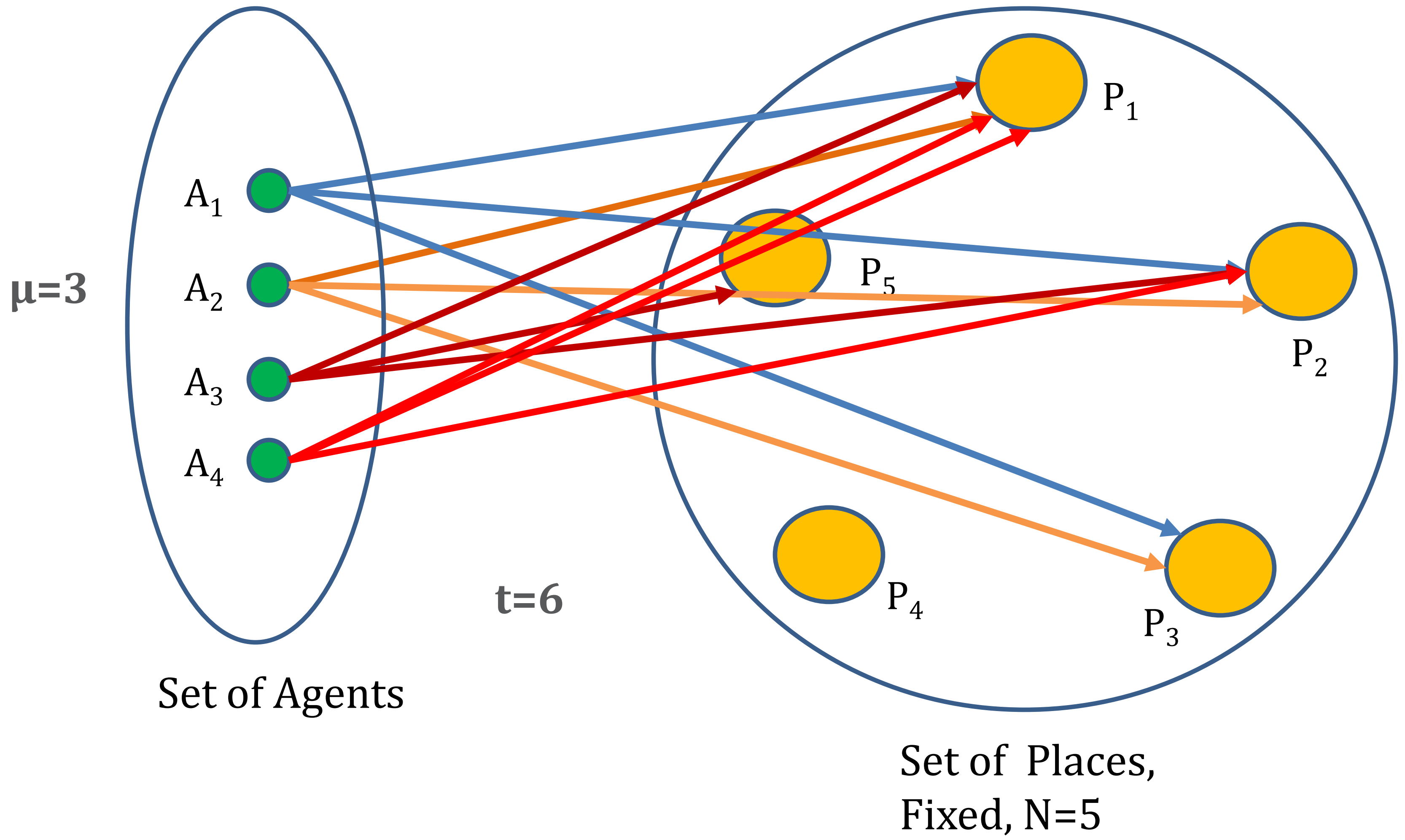
Set of Agents

Set of Places, Fixed, N=5

Overlapping Life Span

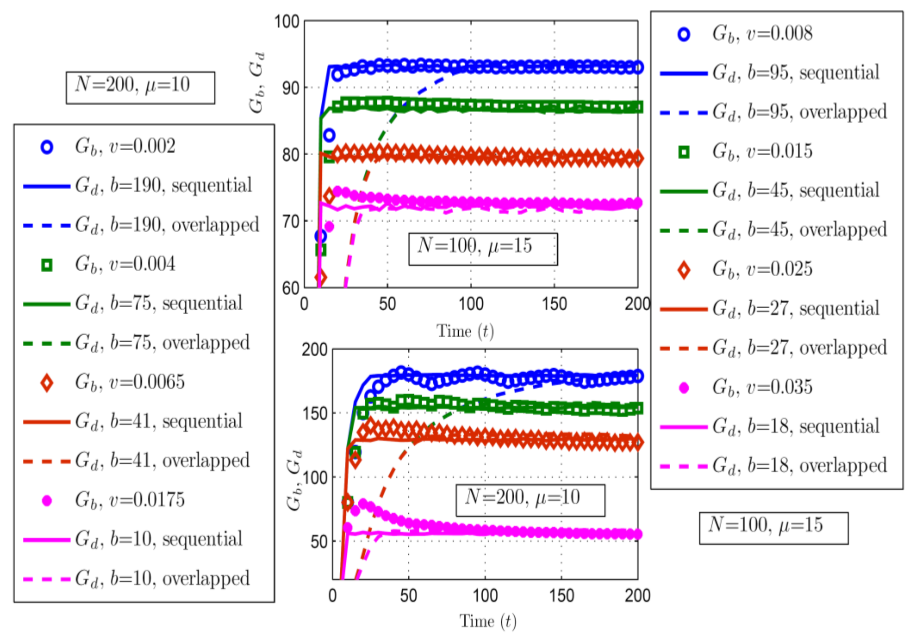


Overlapping Life Span



Overlapping Life Span

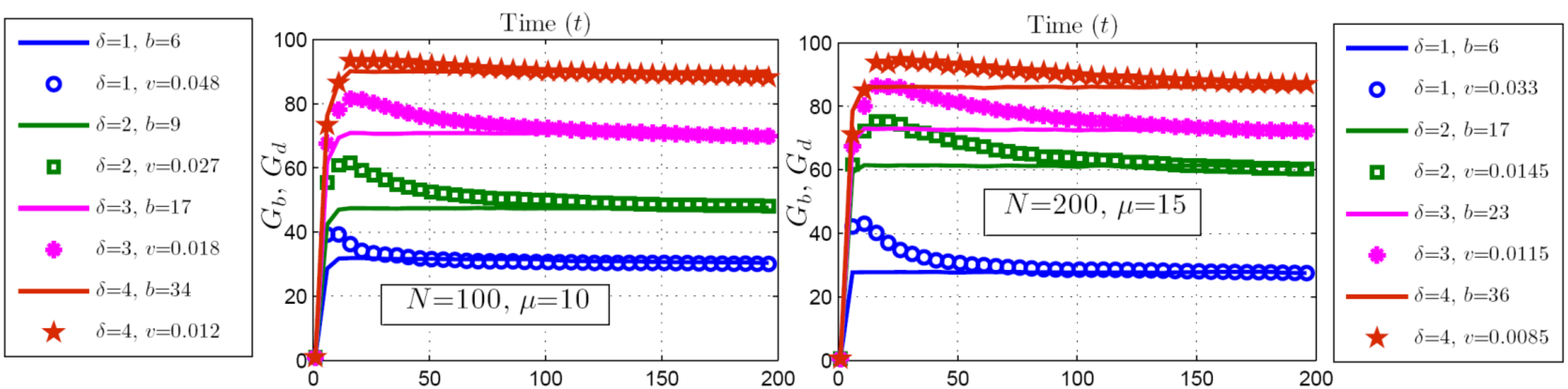
- Redefining the concept of time
  - Globally measured and independent of the number of agents
  - There can be more than one agent in a given time step
  - For simplicity
    - Within a single time step, the agents connect to the places according to their arrival order



- Primary change
  - The agents who come later may not get the full benefit of the spreading done by the agents who joined earlier
  - Stability comes later but the saturation value is same as that of disjoint case. Therefore, bipartite modeling works fine

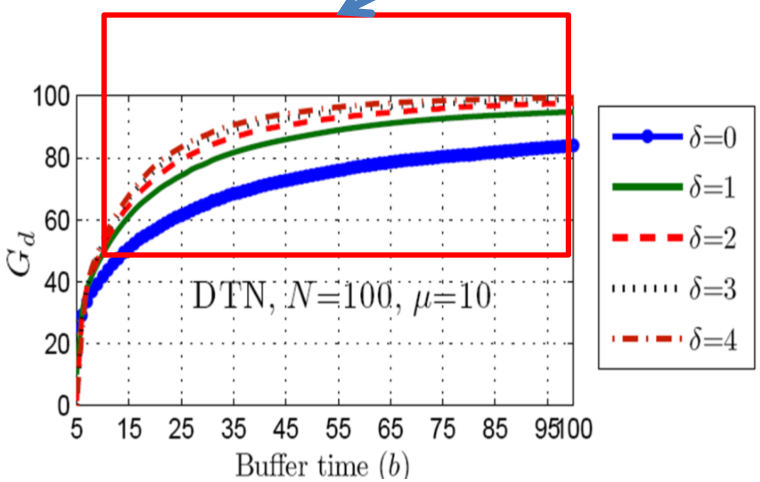
Effect of Randomness

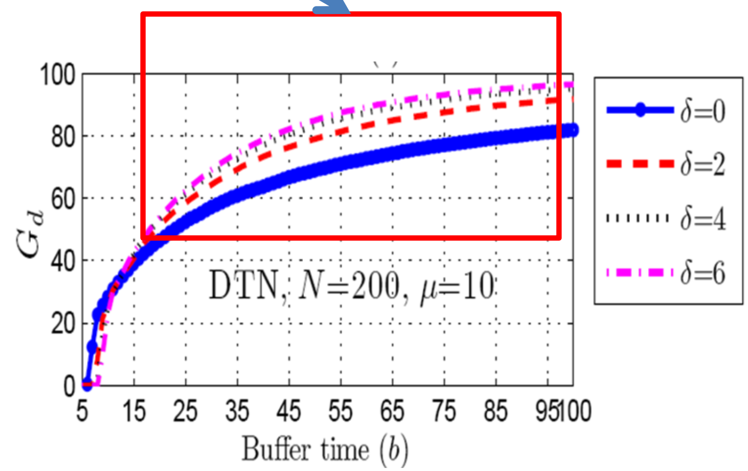
- Randomness in selection of the next place
  - Probability of selecting the i[th] place by an agent is calculated by the formula where $d_i(t)$ is the degree of the place node at time step t, N is the total number of place nodes available and δ is the parameter controlling the randomness in the agents' choice $$\frac{d_i(t)+\delta}{\sum_{j=1}^{N}(d_i(t)+\delta)}$$
  - The time evolution of the coverage and the number of nodes in the largest connected component still match quite well (see following figure)



Effect of Randomness

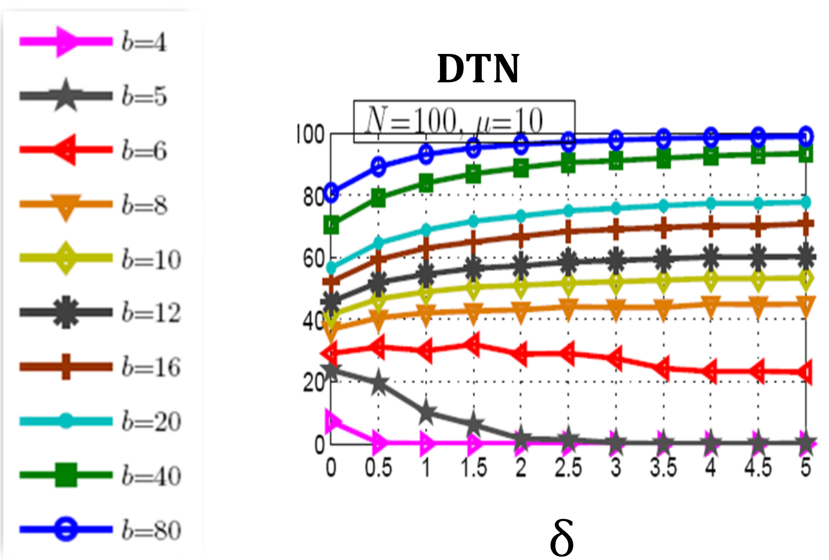
- The critical buffer time depends on
  - Total number of places (N)
  - The average number of connections created by an agent (μ)
- From the design engineer's perspective, the employed buffer time should be always higher than the critical buffer time so that the effect of randomness in the agents' choice is positive, i.e., brings more coverage

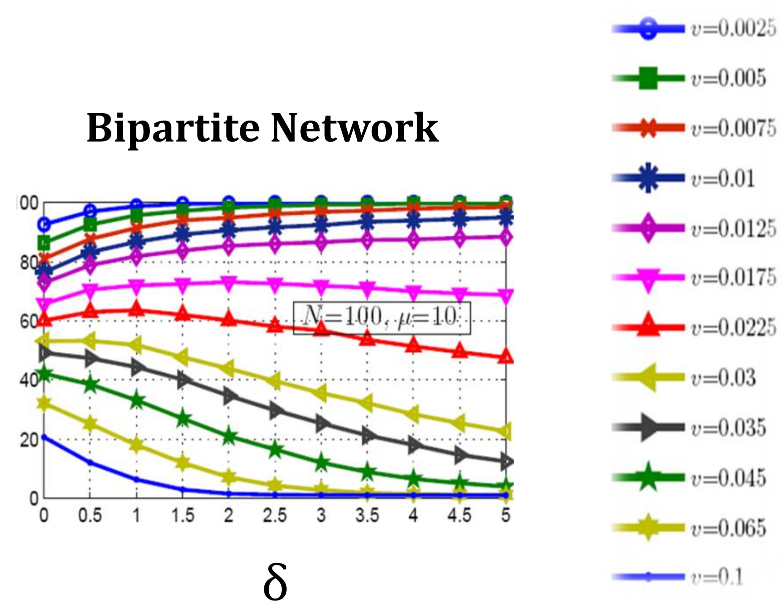




Effect of Randomness

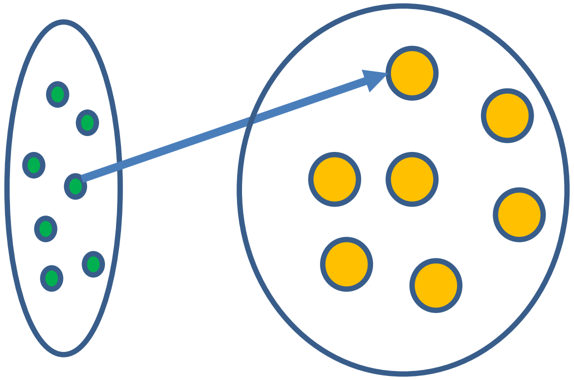
- The critical buffer time depends on
  - Total number of places (N)
  - The average number of connections created an agent (μ)
- From the design engineer's perspective, the employed buffer time should be always higher than the critical buffer time so that the effect of randomness in the agents' choice is positive , i.e., brings more coverage





Summary of Contribution

- Correlation of DTN with Bipartite network
  - Information dissemination in DTN and the evolution of the largest component size in one-mode projection of Bipartite Network
- Mathematical analysis of achieved coverage under a given buffer time in DTN
- Relationship between the average social participation of the agents, total number of places with the achieved coverage
- Notion of optimal buffer time
- Analysis of the effect of overlapping life spans
- Analysis of the effect of randomness



Future Work

- Simplification of the relation between v and b
- Full exploration of the correlation between the effect of randomness
  - Deducing the value of the time varying threshold (v) for a given buffer time (b) under a certain value of the randomness parameter (δ)
  - Exploring the status of the special property (P) under randomness
  - Estimating the coverage in DTN under a given randomness
- Testing the theoretical results under real data
- Reformulation using
  - Variable buffer time
  - Variable popularity of a given piece of information

Publications

❑ **Publication out of this work :**

1. Understanding Information Dissemination Dynamics in Delay Tolerant Network using Theory of Bipartite Network. Sudipta Saha, Niloy Ganguly and Animesh Mukherjee. PhD Forum, COMSNETS 2012 (Got the best PhD forum presentation award)

2. Information Dissemination Dynamics in Delay Tolerant Network: A Bipartite Network Approach. Sudipta Saha, Niloy Ganguly and Animesh Mukherjee. In Proceedings of ACM MobiOpp, 2012

❑ **Interested readers may visit the following site for more details on the topic** **http://cse.iitkgp.ac.in/~sudiptas/dtnbnw.html**

A Special Thanks to TCS for the Financial Support

Thank you all for the patience

Sudipta Saha

sudipta.saha@cse.iitkgp.ernet.in

http://cse.iitkgp.ac.in/~sudiptas

TCS Research Scholar Program