Networks of Information Exchange: Evidence on Information Hubs

Pritha Dev
Indian Institute of Management Ahmedabad

March 14, 2016

Introduction

Introduction

- This paper presents a empirical investigation to tests the prediction of hubs

Introduction

- This paper presents a empirical investigation to tests the prediction of hubs
  - By checking if the probability of link formation with a player is increasing in the number of other links the player has.

Introduction

- This paper presents a empirical investigation to tests the prediction of hubs
  - By checking if the probability of link formation with a player is increasing in the number of other links the player has.
  - It proposes a novel way to measure the number of links of the match

Introduction

- This paper presents a empirical investigation to tests the prediction of hubs
  - By checking if the probability of link formation with a player is increasing in the number of other links the player has.
  - It proposes a novel way to measure the number of links of the match
  - And controls for the endogeneity of this independent variable.

Related Literature

Related Literature

- The theoretical literature in economics on network formation follows two main strands - one follows Jackson and Wollinsky (1996) and the other follows Bala and Goyal (2000).

Related Literature

- The theoretical literature in economics on network formation follows two main strands - one follows Jackson and Wollinsky (1996) and the other follows Bala and Goyal (2000).
- Galeotti and Goyal (2010) consider production of information on a network. Hojman and Szeidl (2008) and Bloch and Dutta (2009) show periphery sponsored stars arising in an information network.

Related Literature

- The theoretical literature in economics on network formation follows two main strands - one follows Jackson and Wollinsky (1996) and the other follows Bala and Goyal (2000).
- Galeotti and Goyal (2010) consider production of information on a network. Hojman and Szeidl (2008) and Bloch and Dutta (2009) show periphery sponsored stars arising in an information network.
- Empirical papers investigating the formation of networks includes work by Udry and Conley(2004), Santos and Barrett(2004), Fafchamps and Lund (2003), De Weerdt (2004) and others. Links are found to be based on kinship, religion, common friends, etc.

Related Literature

- The theoretical literature in economics on network formation follows two main strands - one follows Jackson and Wollinsky (1996) and the other follows Bala and Goyal (2000).
- Galeotti and Goyal (2010) consider production of information on a network. Hojman and Szeidl (2008) and Bloch and Dutta (2009) show periphery sponsored stars arising in an information network.
- Empirical papers investigating the formation of networks includes work by Udry and Conley(2004), Santos and Barrett(2004), Fafchamps and Lund (2003), De Weerdt (2004) and others. Links are found to be based on kinship, religion, common friends, etc.
- Comola (2007) looks at the models bilateral link formation and the impact of links of links. The value of each link is measured by wealth.

Networks of Information Exchange: Theoretical Basis

Networks Formation Games: Star Networks

Networks Formation Games: Star Networks

- A star network is such that there is a cental player and all other players form links with this central player and no other links.

Networks Formation Games: Star Networks

- A star network is such that there is a cental player and all other players form links with this central player and no other links.
- Star networks arise if

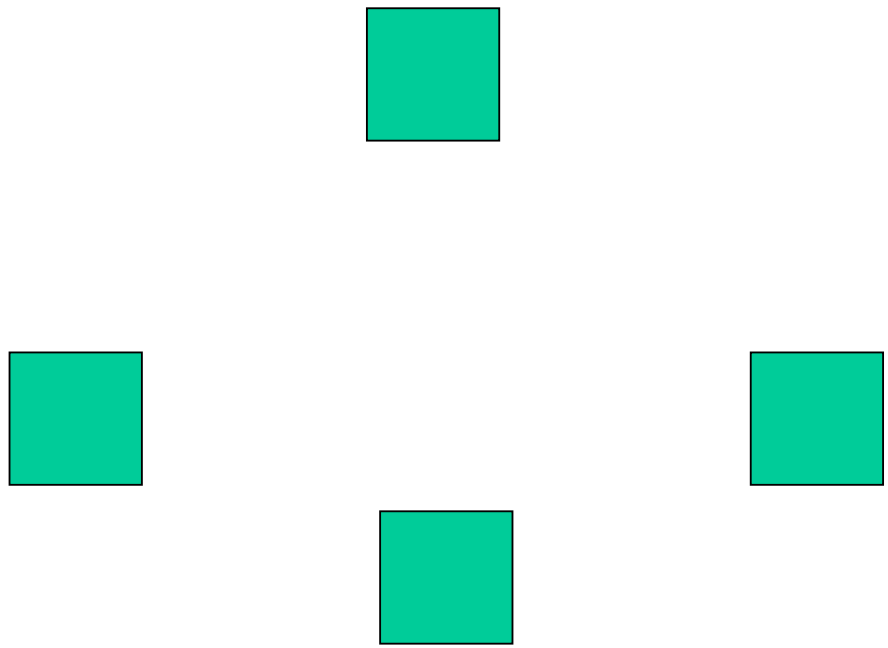
Networks Formation Games: Star Networks

- A star network is such that there is a cental player and all other players form links with this central player and no other links.
- Star networks arise if
  - Information is not passed along perfectly but with some **decay**

Networks Formation Games: Star Networks

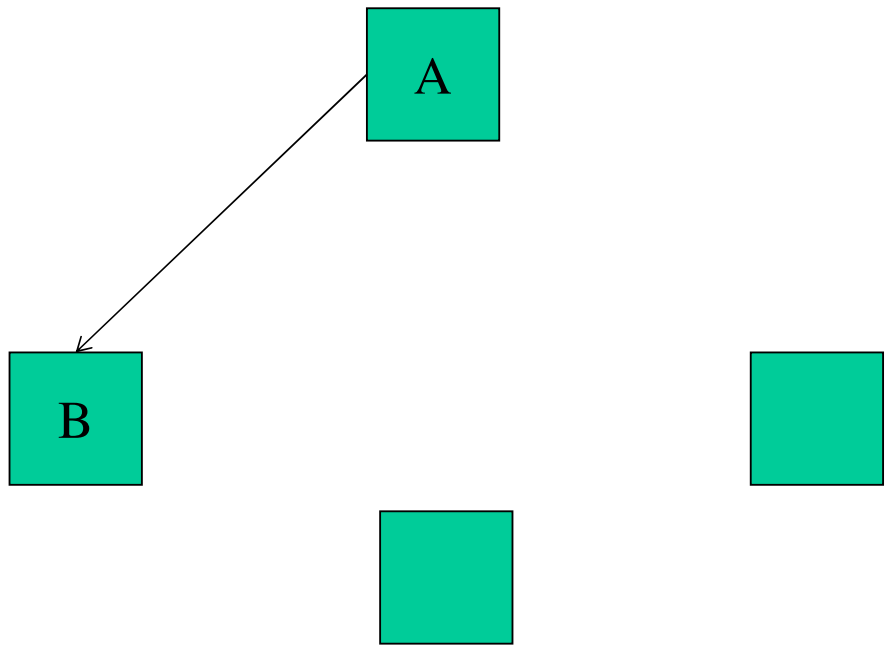
- A star network is such that there is a cental player and all other players form links with this central player and no other links.
- Star networks arise if
  - Information is not passed along perfectly but with some **decay**
  - The **cost** of link formation is high enough

Network Formation: Role of Decay

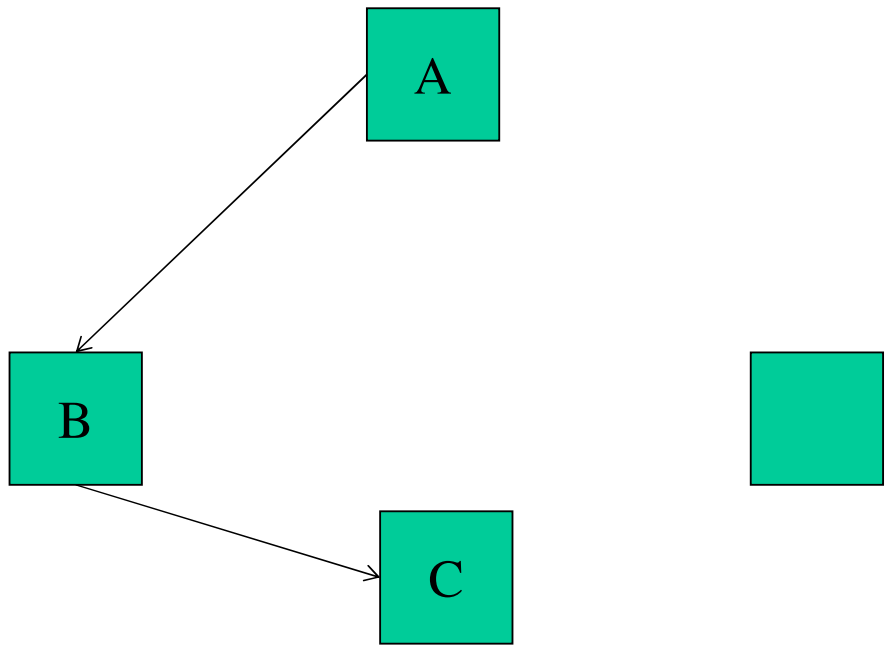


4 node network

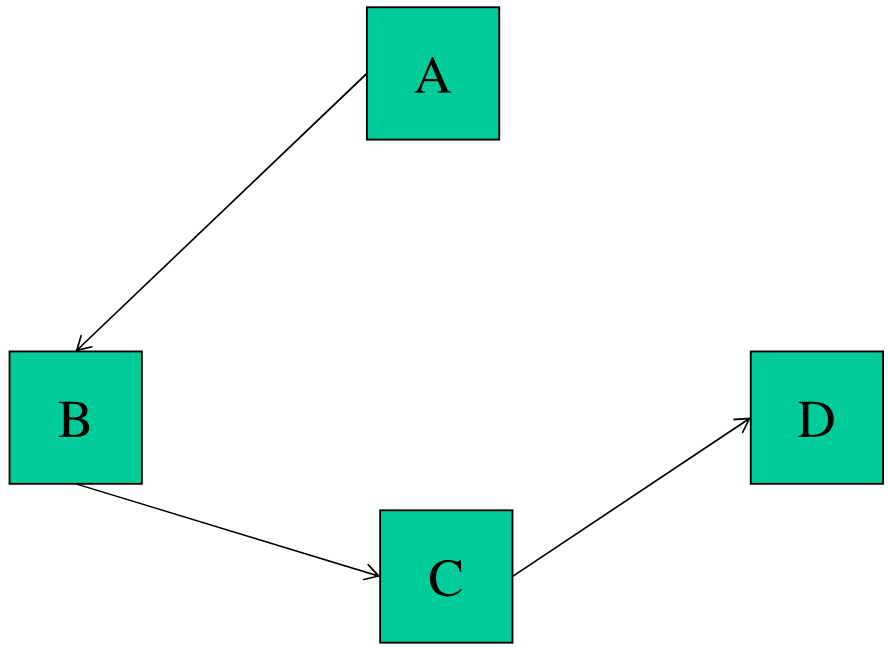
Any Network Possible Without Decay



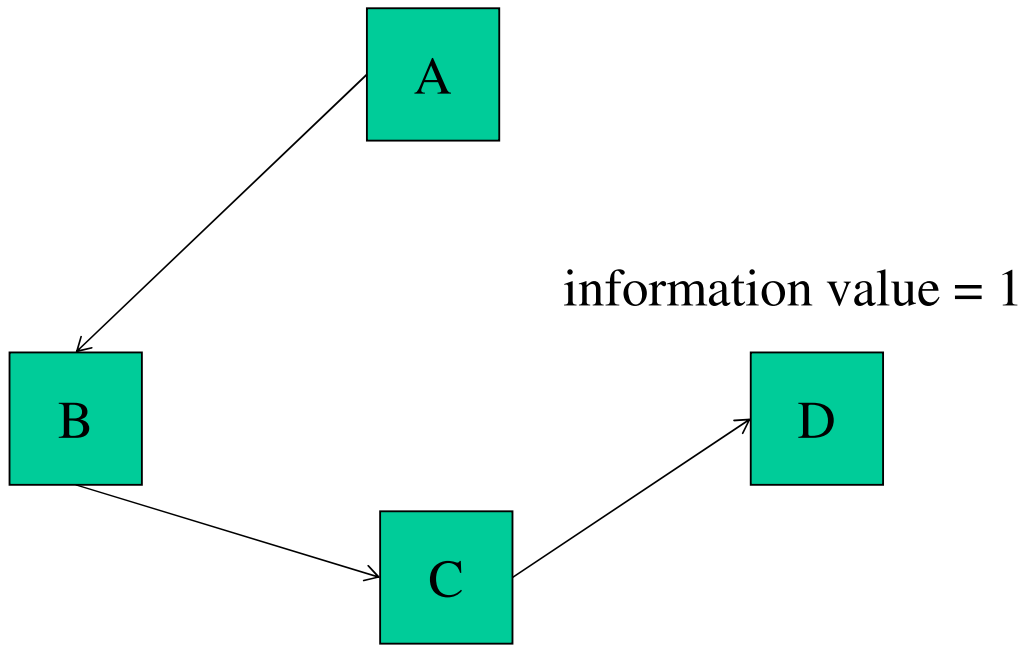
Any Network Possible Without Decay



Any Network Possible Without Decay

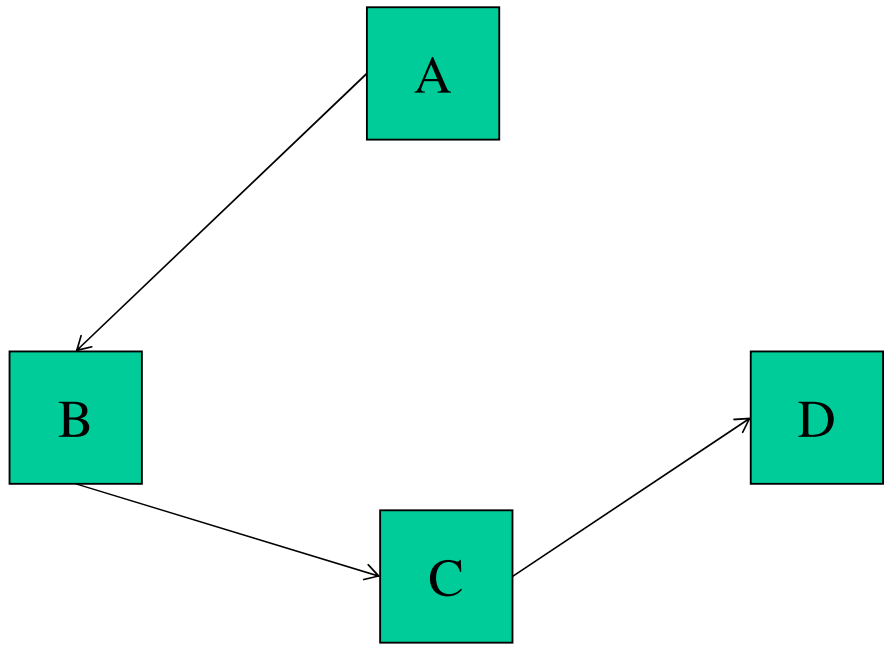


Information Flow is Perfect Without Decay



Information Flow With NO Decay

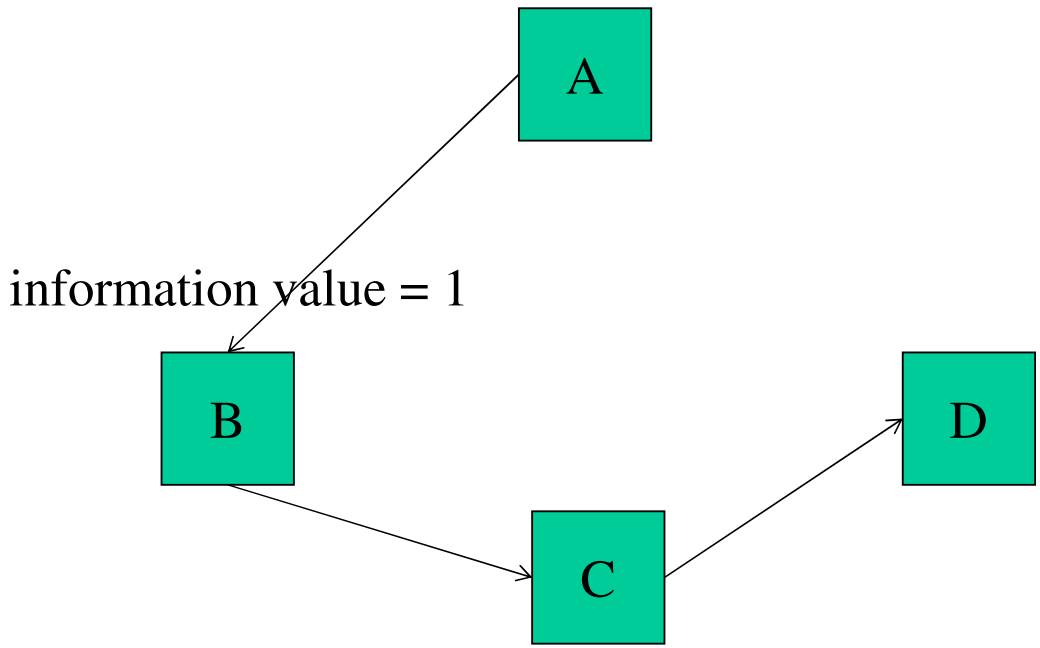
Information Flow is Perfect Without Decay



information value = 1

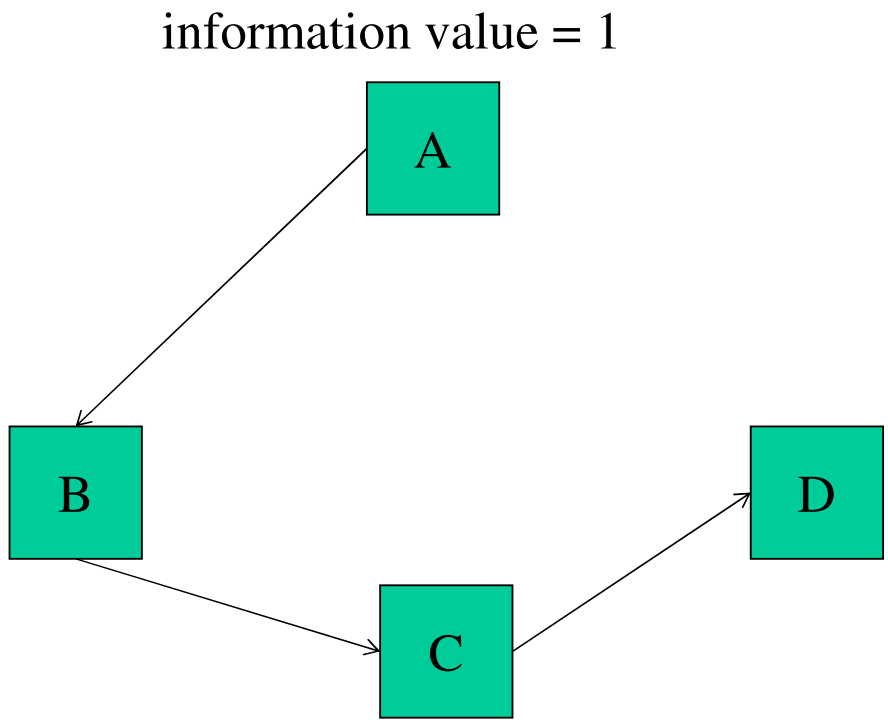
Information Flow With NO Decay

Information Flow is Perfect Without Decay



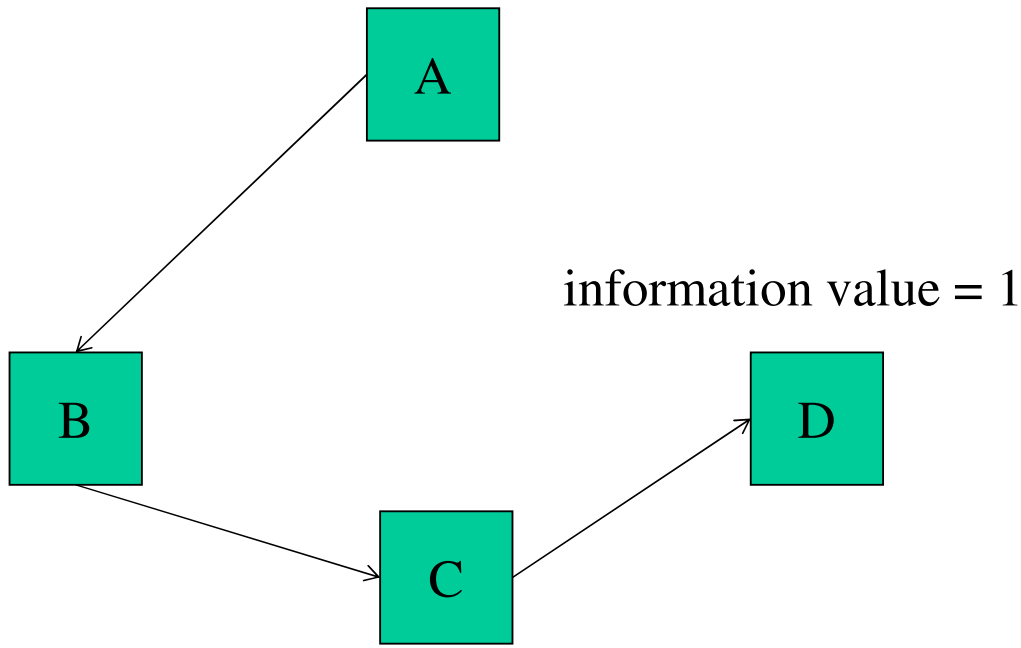
Information Flow With NO Decay

Information Flow is Perfect Without Decay



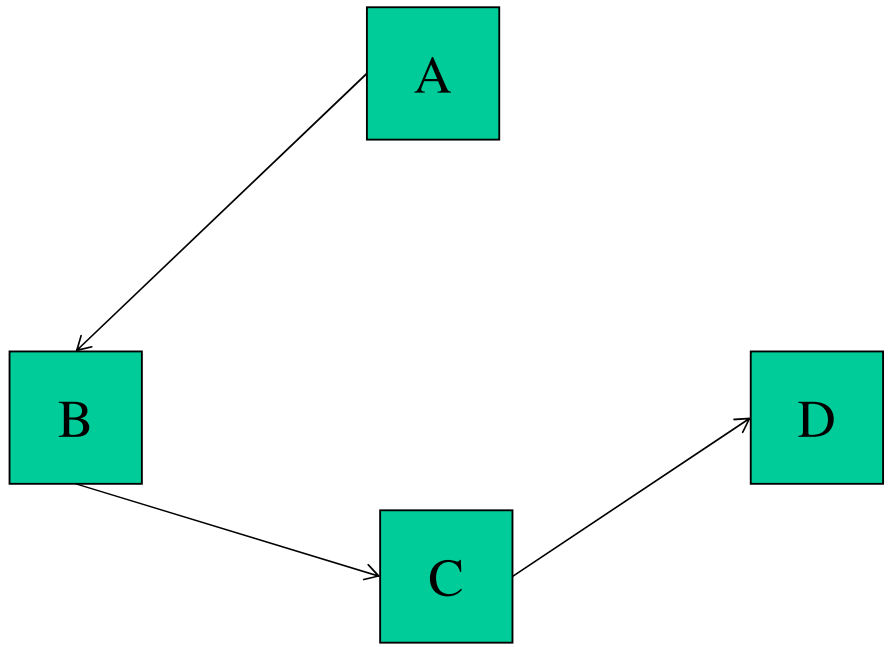
Information Flow With NO Decay

Information Flow is NOT Perfect With Decay



Information Flow With Decay of 10%

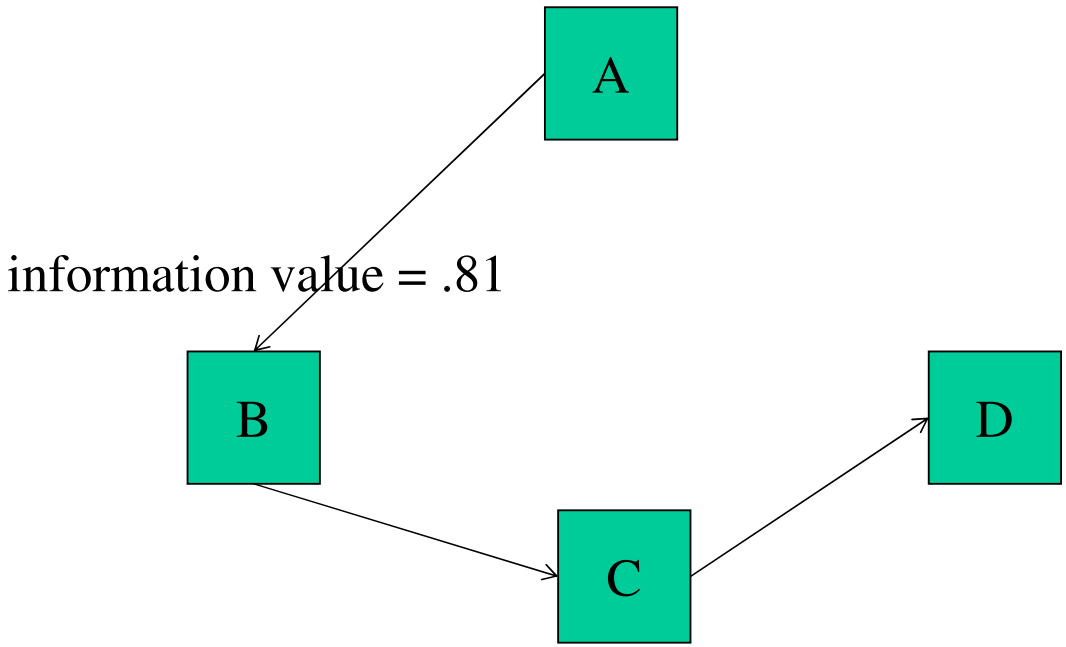
Information Flow is NOT Perfect With Decay



information value = .9

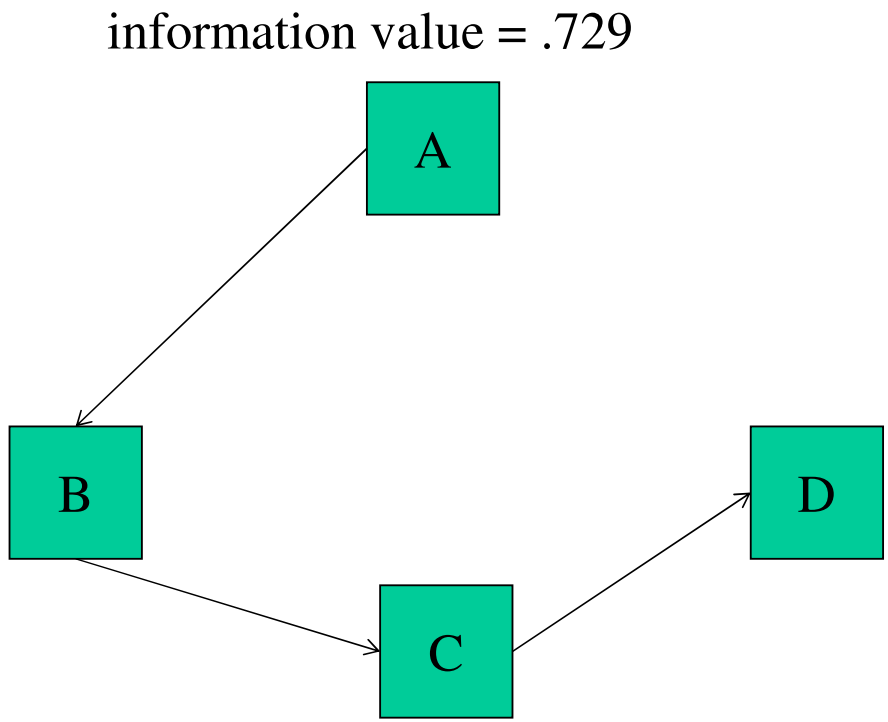
Information Flow With Decay of 10%

Information Flow is NOT Perfect With Decay



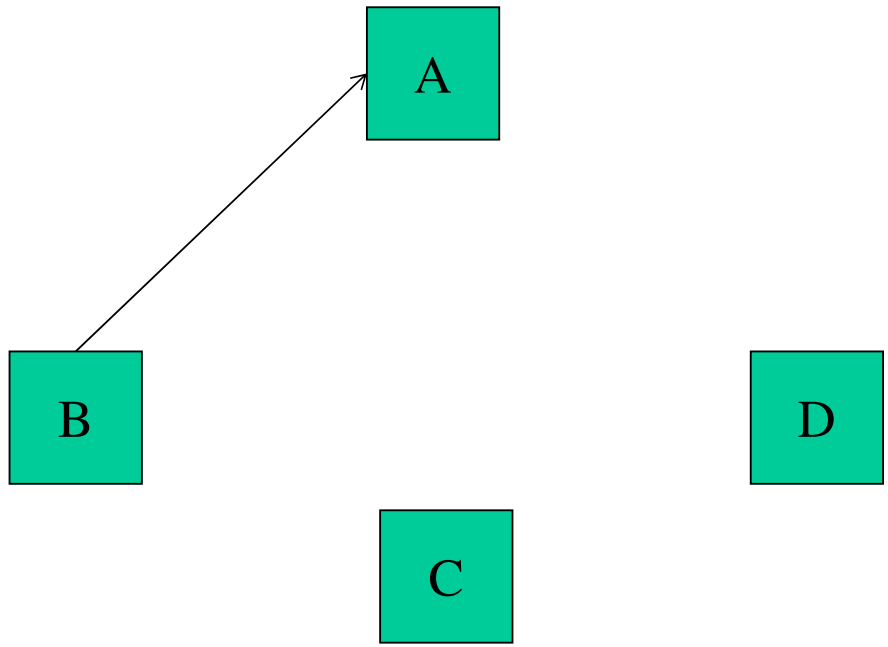
Information Flow With Decay of 10%

Information Flow is NOT Perfect With Decay

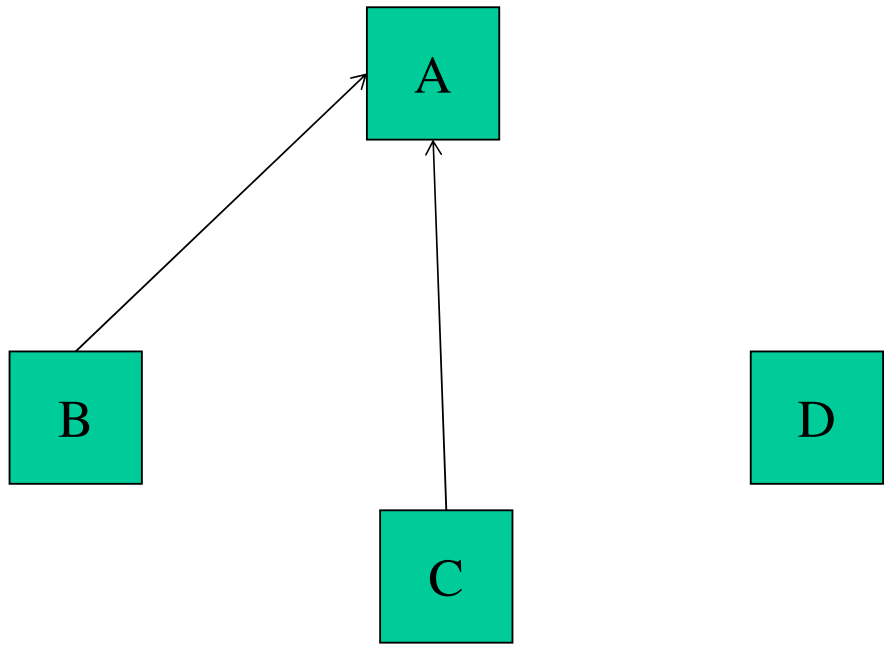


Information Flow With Decay of 10%

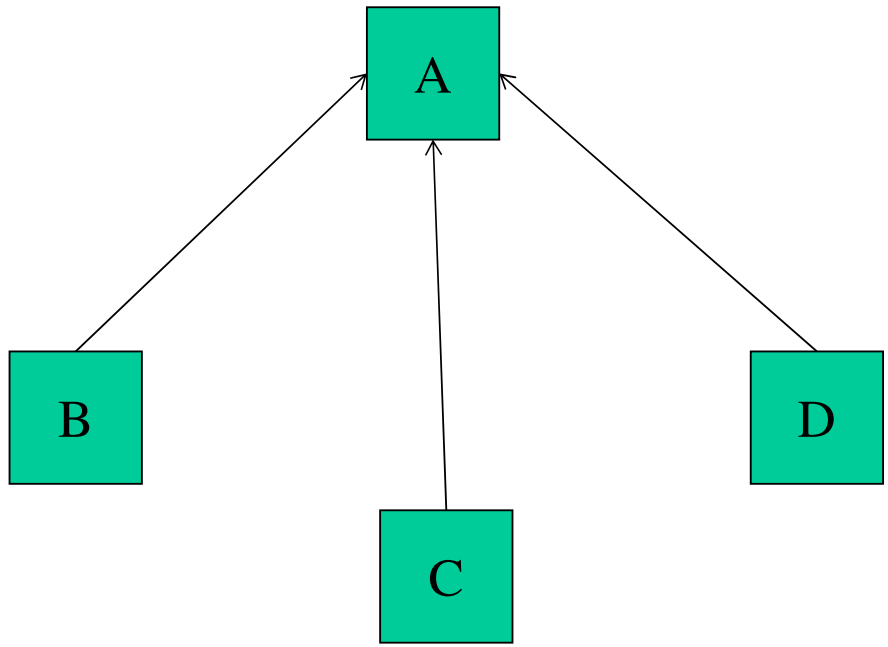
Star Network: Most Efficient With Decay



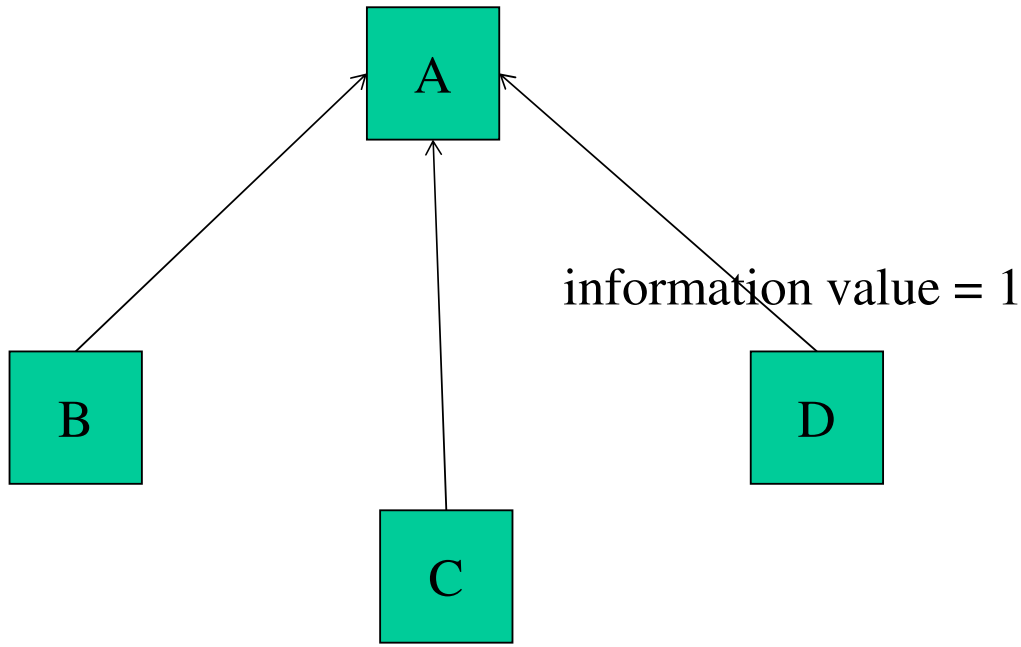
Star Network: Most Efficient With Decay



Star Network: Most Efficient With Decay

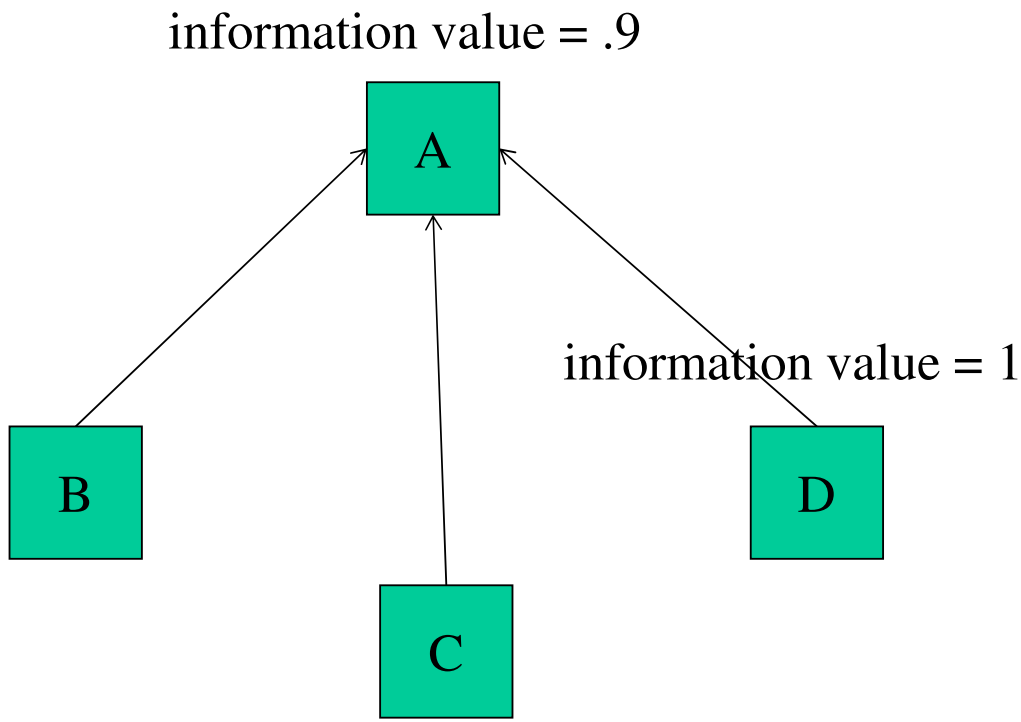


Star Network: Most Efficient With Decay



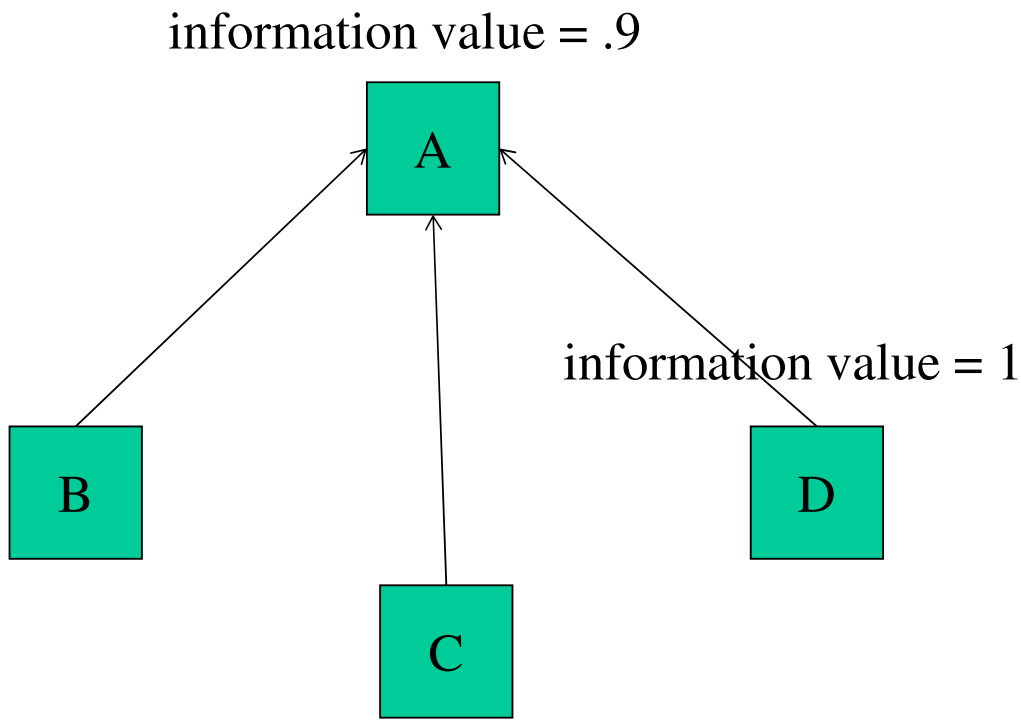
Information Flow With Decay of 10%

Star Network: Most Efficient With Decay



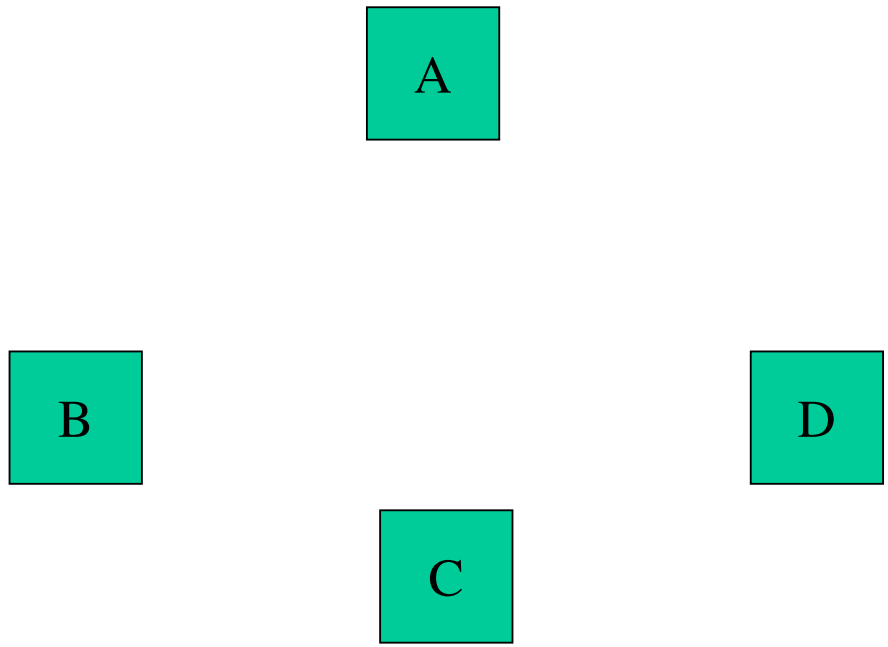
Information Flow With Decay of 10%

Star Network: Most Efficient With Decay



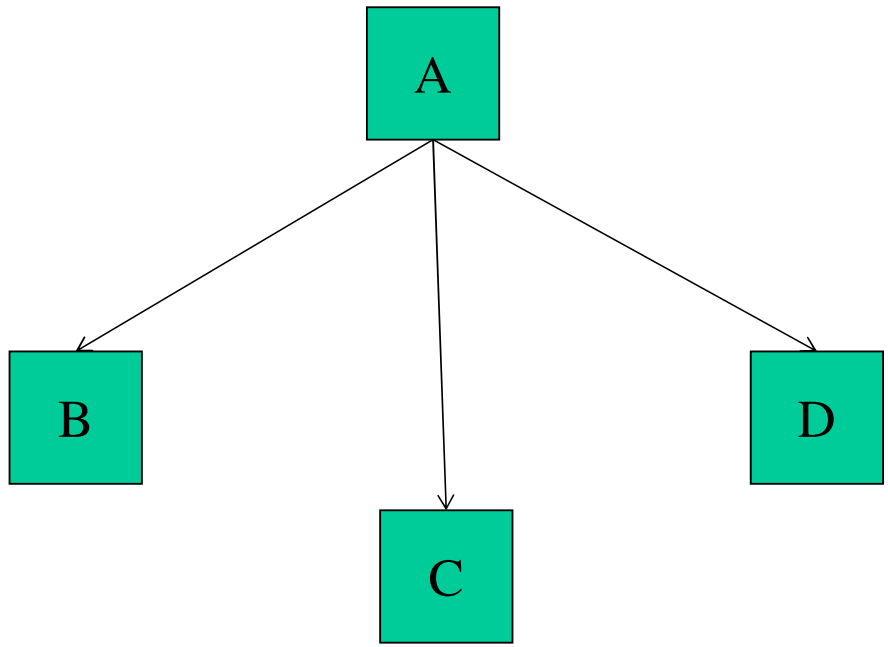
Information Flow With Decay of 10%

Network Formation: Role of Link Cost



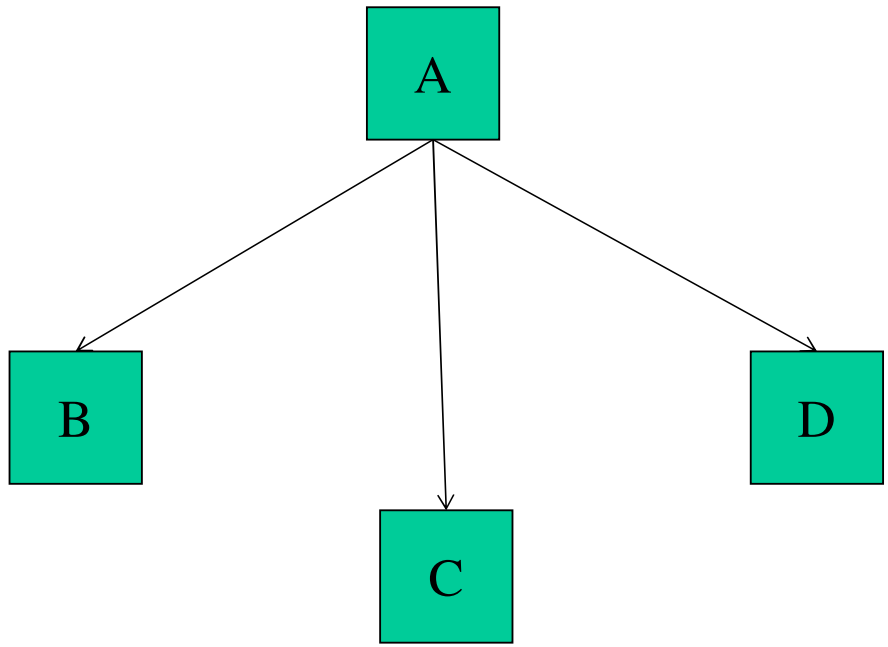
Network Formation with NO Cost

Network Formation: Role of Link Cost



Network Formation with NO Cost

Network Formation: Role of Link Cost

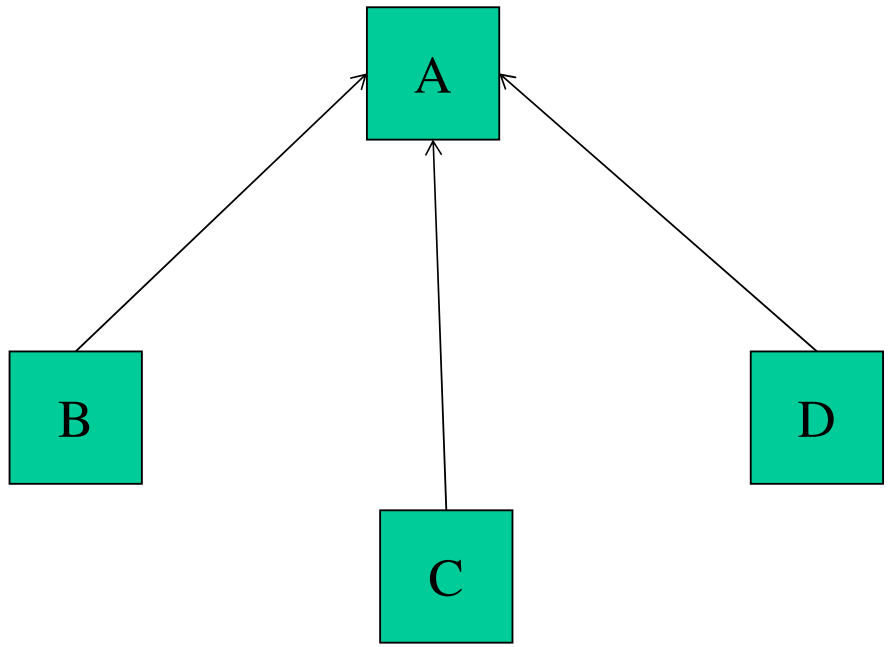


Network Formation with NO Cost

=>

Everyone Links to Everyone Else

Decay + High Cost = Star Network



Information Aggregator

Information Aggregator

- In the above game, *any* player can end up being a hub who receives at least three links.

Information Aggregator

- In the above game, *any* player can end up being a hub who receives at least three links.
- An **Information Aggregator** is a player who can double (or more) the value of any information received by him and transmit it (back) to his direct links.

Information Aggregator

- In the above game, *any* player can end up being a hub who receives at least three links.
- An **Information Aggregator** is a player who can double (or more) the value of any information received by him and transmit it (back) to his direct links.
- If the above game is changed to have one player who is an Information Aggregator, then any non-empty Nash network has the structure of a periphery sponsored star with this player as the center.

Networks Formation Games: Interconnected Star Networks

Networks Formation Games: Interconnected Star Networks

- Instead of the network being represented by a single star, interconnected star networks are commonly observed.

Networks Formation Games: Interconnected Star Networks

- Instead of the network being represented by a single star, interconnected star networks are commonly observed.
- Some players are local hubs of information

Networks Formation Games: Interconnected Star Networks

- Instead of the network being represented by a single star, interconnected star networks are commonly observed.
- Some players are local hubs of information
- Local hubs are then connected to each other

Networks Formation Games: Interconnected Star Networks

- Instead of the network being represented by a single star, interconnected star networks are commonly observed.
- Some players are local hubs of information
- Local hubs are then connected to each other
- Such networks arise when costs of link formation are variable and depend on group membership/social distance.

Networks of Information Exchange: Empirics

The Empirical Model

The Empirical Model

- According to theoretical results, the probability of forming a link by a respondent to a match is a function of:

The Empirical Model

- According to theoretical results, the probability of forming a link by a respondent to a match is a function of:
  - the total links of the match

The Empirical Model

- According to theoretical results, the probability of forming a link by a respondent to a match is a function of:
  - the total links of the match
  - individual characteristics of the match capturing the value of his information/level of information aggregation

The Empirical Model

- According to theoretical results, the probability of forming a link by a respondent to a match is a function of:
  - the total links of the match
  - individual characteristics of the match capturing the value of his information/level of information aggregation
  - the social distance between the respondent and match capturing the cost of the link

The Empirical Model: Issues

The Empirical Model: Issues

- The independent variable of interest is the total links of the match.

The Empirical Model: Issues

- The independent variable of interest is the total links of the match.
- There is **measurement bias** in this variable

The Empirical Model: Issues

- The independent variable of interest is the total links of the match.
- There is **measurement bias** in this variable
  - If the entire network is not observed, this variable is mismeasured.

The Empirical Model: Issues

- The independent variable of interest is the total links of the match.
- There is **measurement bias** in this variable
  - If the entire network is not observed, this variable is mismeasured.
  - Moreover, the variable is measured below its true value.

The Empirical Model: Issues

- The independent variable of interest is the total links of the match.
- There is **measurement bias** in this variable
  - If the entire network is not observed, this variable is mismeasured.
  - Moreover, the variable is measured below its true value.
- This variable is also **endogenous.**

The Empirical Model: Issues

- The independent variable of interest is the total links of the match.
- There is **measurement bias** in this variable
  - If the entire network is not observed, this variable is mismeasured.
  - Moreover, the variable is measured below its true value.
- This variable is also **endogenous.**
  - This is so since in theory, all link decisions are made strategically and simultaneously.

The Empirical Model: Issues

- The independent variable of interest is the total links of the match.
- There is **measurement bias** in this variable
  - If the entire network is not observed, this variable is mismeasured.
  - Moreover, the variable is measured below its true value.
- This variable is also **endogenous.**
  - This is so since in theory, all link decisions are made strategically and simultaneously.
  - The total links of the match is just the sum of many such decisions.

Measurement Bias: Problem

Measurement Bias: Problem

- The network data collected by first selecting a random sample of persons/nodes

Measurement Bias: Problem

- The network data collected by first selecting a random sample of persons/nodes
- Within the selected nodes, each person is asked whether he links or not with a random sample from the remaining nodes

Measurement Bias: Problem

- The network data collected by first selecting a random sample of persons/nodes
- Within the selected nodes, each person is asked whether he links or not with a random sample from the remaining nodes
- Hence, it must be that we NEVER observe the full number of links any player has

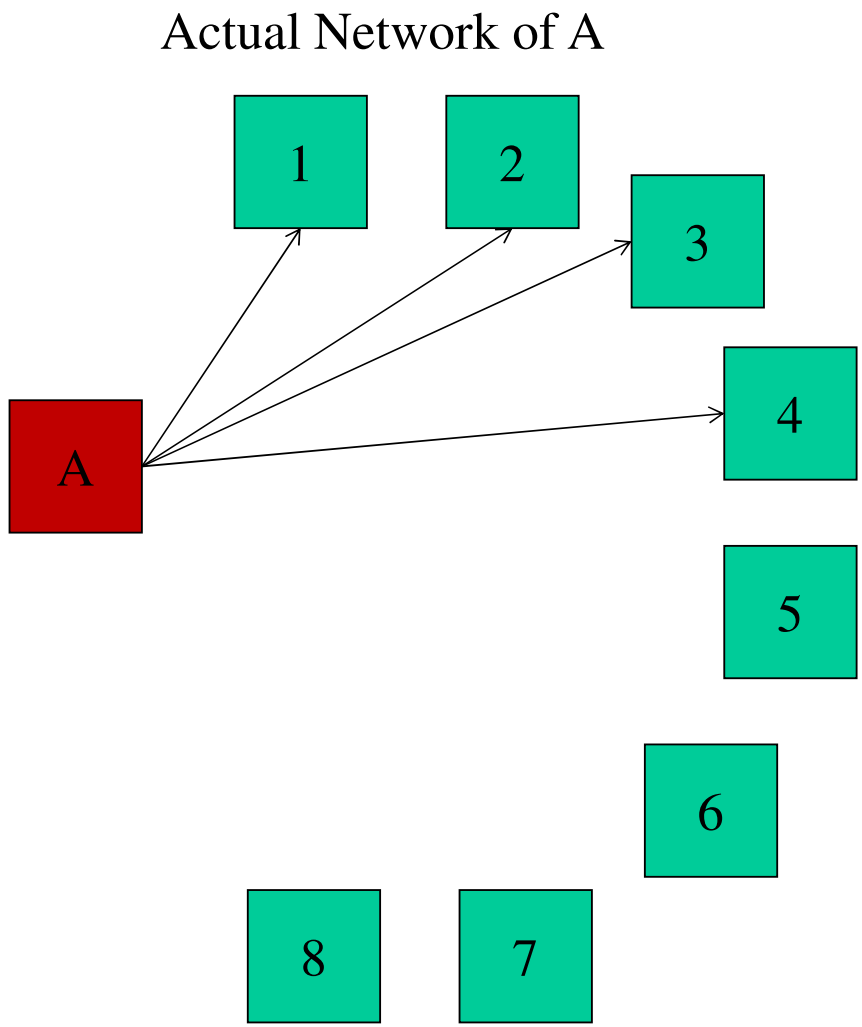
Measurement Bias: Problem

- The network data collected by first selecting a random sample of persons/nodes
- Within the selected nodes, each person is asked whether he links or not with a random sample from the remaining nodes
- Hence, it must be that we NEVER observe the full number of links any player has
- In fact, we always observe LESS than the total links made by a player

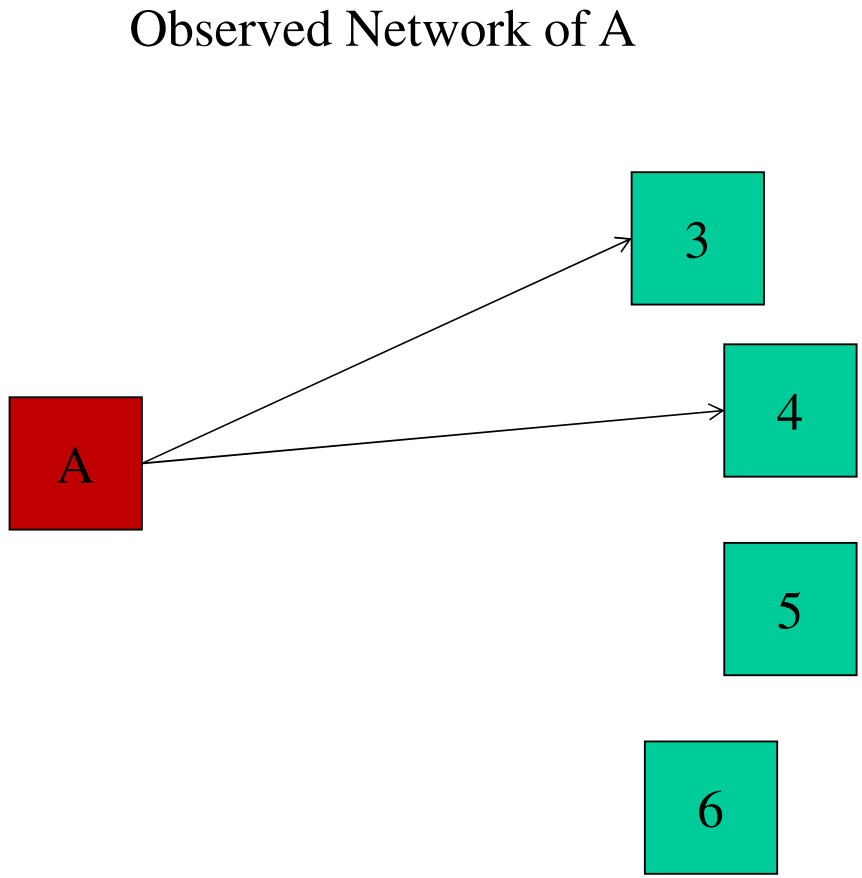
Measurement Bias: Problem

- The network data collected by first selecting a random sample of persons/nodes
- Within the selected nodes, each person is asked whether he links or not with a random sample from the remaining nodes
- Hence, it must be that we NEVER observe the full number of links any player has
- In fact, we always observe LESS than the total links made by a player
- Or, the bias is always negative.

Measurement Bias: Problem

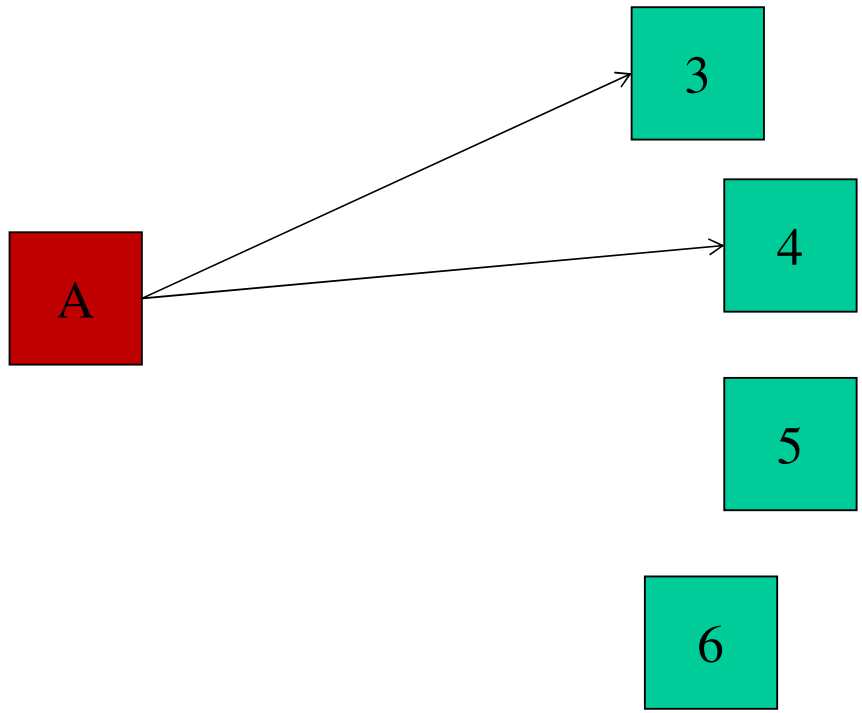


Measurement Bias: Problem



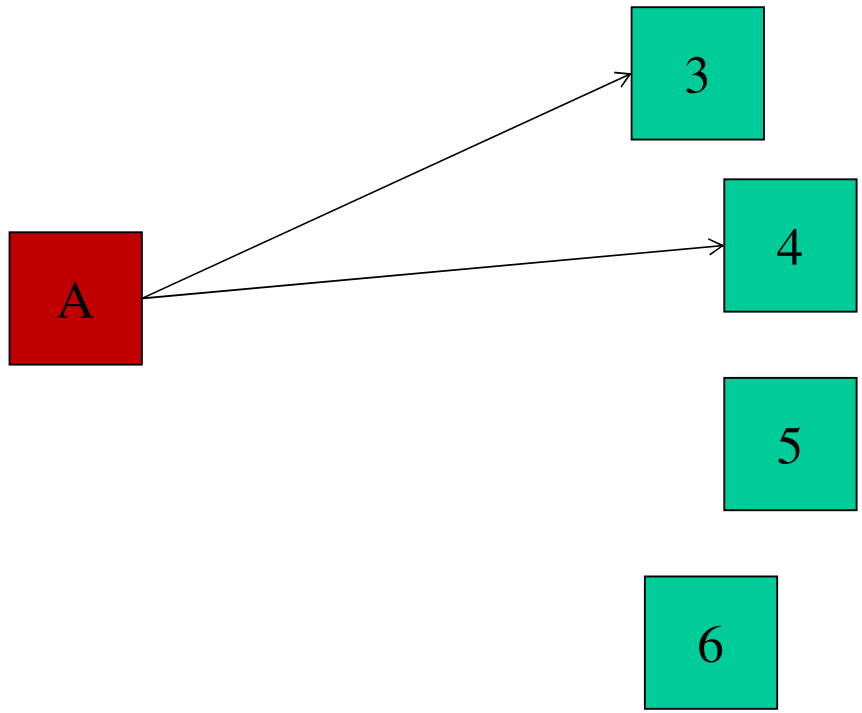
Measurement Bias: Problem

Observed Links < Actual Links



Measurement Bias: Problem

Observed Links < Actual Links
=> Measurement Bias



Measurement Bias: Solution

Measurement Bias: Solution

- Consider the **proportion of links received** by a node, measured as the actual links received by the node in the sample divided by the total number of other nodes who were asked if they have links with the node.

Measurement Bias: Solution

- Consider the **proportion of links received** by a node, measured as the actual links received by the node in the sample divided by the total number of other nodes who were asked if they have links with the node.
- The proportion of links made by a node can be similarly measured.

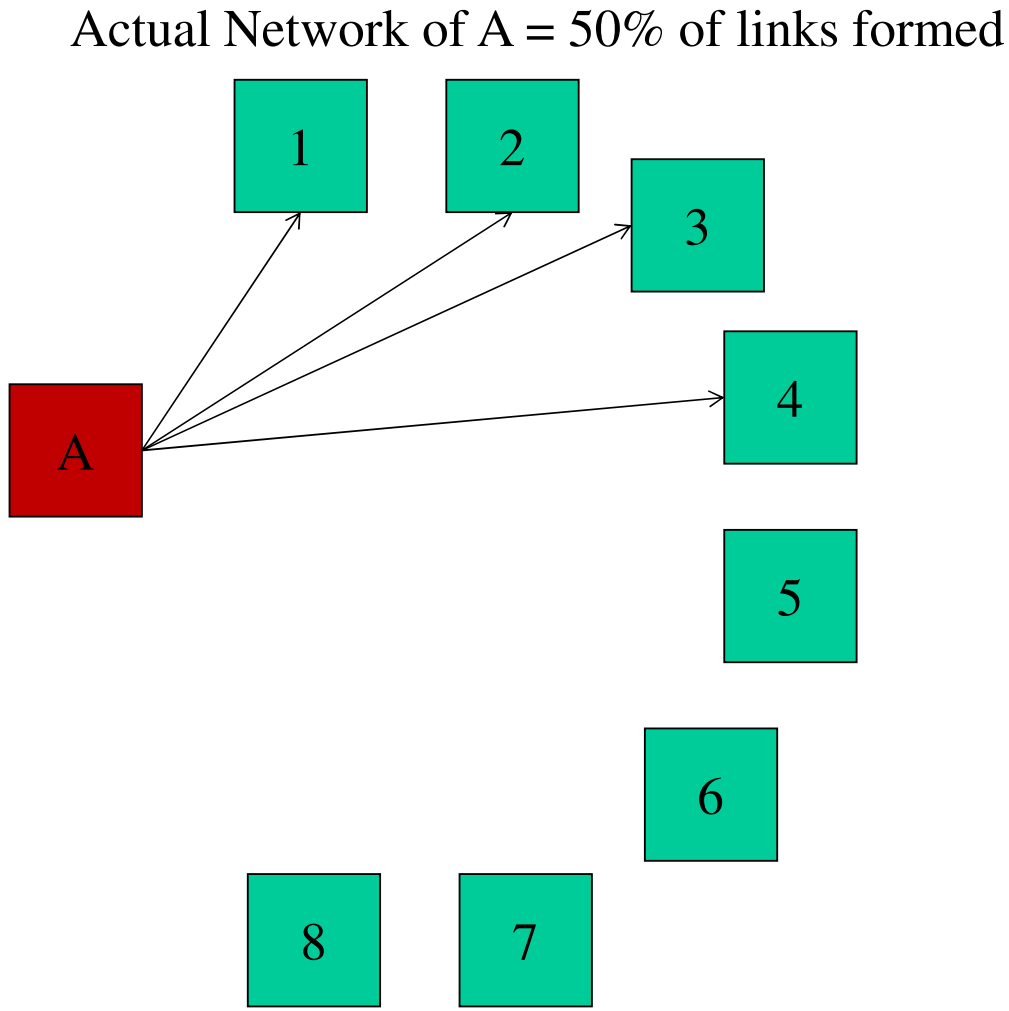
Measurement Bias: Solution

- Consider the **proportion of links received** by a node, measured as the actual links received by the node in the sample divided by the total number of other nodes who were asked if they have links with the node.
- The proportion of links made by a node can be similarly measured.
- While the total links received or made by a node are measured with a negative bias, the proportion of links received/made is not

Measurement Bias: Solution

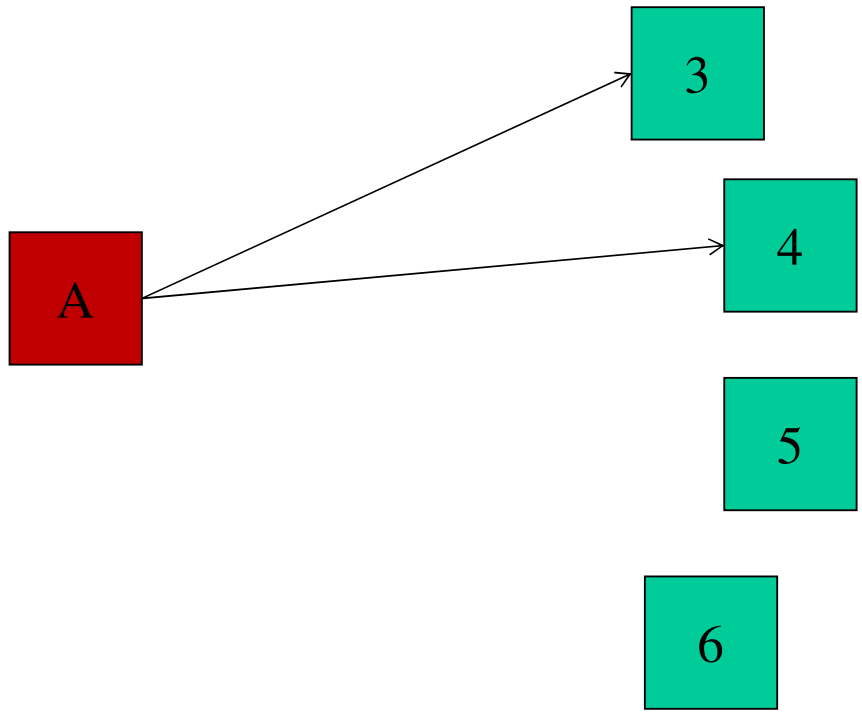
- Consider the **proportion of links received** by a node, measured as the actual links received by the node in the sample divided by the total number of other nodes who were asked if they have links with the node.
- The proportion of links made by a node can be similarly measured.
- While the total links received or made by a node are measured with a negative bias, the proportion of links received/made is not
- Hence, the mismeasured variable is replaced by the variable measuring the proportion of links received/made by a node which is directly related to the mismeasured variable.

Measurement Bias: Solution



Measurement Bias: Problem

Observed Network of A = 50% of links formed



Endogeneity: Problem

Endogeneity: Problem

- This paper seeks to measure if the A's decision to form a link with B depends on the total links of B

Endogeneity: Problem

- This paper seeks to measure if the A's decision to form a link with B depends on the total links of B
- The independent variable is the total links of B, which depend on the decisions to link taken by B.

Endogeneity: Problem

- This paper seeks to measure if the A's decision to form a link with B depends on the total links of B
- The independent variable is the total links of B, which depend on the decisions to link taken by B.
- Since all the decisions to link are assumed to be taken simultaneously, there is the concern of endogeneity.

Endogeneity: Solution

Endogeneity: Solution

- We use the control function approach, where as suggested by the theoretical model, the total links of player B are modeled such that:

Endogeneity: Solution

- We use the control function approach, where as suggested by the theoretical model, the total links of player B are modeled such that:
  - The total links a player receives/makes depend on the average cost of linking to him and level of information aggregation

Endogeneity: Solution

- We use the control function approach, where as suggested by the theoretical model, the total links of player B are modeled such that:
  - The total links a player receives/makes depend on the average cost of linking to him and level of information aggregation
  - Or total links of the match can be modeled as the social distance between the match and the representative/average individual as well his level of information aggregation.

Endogeneity: Solution

- We use the control function approach, where as suggested by the theoretical model, the total links of player B are modeled such that:
  - The total links a player receives/makes depend on the average cost of linking to him and level of information aggregation
  - Or total links of the match can be modeled as the social distance between the match and the representative/average individual as well his level of information aggregation.
- The endogeneity is modeled as the correlation between the error terms - the first coming from the estimation of A's decision to link with B and the second coming from the estimations of total links of B

Empirical Model

Empirical Model

- The decision by i to form a link with j is captured by $g_{i,j}$

Empirical Model

- The decision by i to form a link with j is captured by $g_{i,j}$
- Let m_j be the total proportion of links (made or received or both) of player j

Empirical Model

- The decision by i to form a link with j is captured by $g_{i,j}$
- Let m_j be the total proportion of links (made or received or both) of player j
- Let v_j measure the value of player j's information aggregation

Empirical Model

- The decision by i to form a link with j is captured by $g_{i,j}$
- Let m_j be the total proportion of links (made or received or both) of player j
- Let v_j measure the value of player j's information aggregation
- Let X_i, X_j collect individual characteristics of players i, j. Further, let the the average individual's characteristics be denoted by X_A.

Empirical Model

- The decision by i to form a link with j is captured by $g_{i,j}$
- Let m_j be the total proportion of links (made or received or both) of player j
- Let v_j measure the value of player j's information aggregation
- Let X_i, X_j collect individual characteristics of players i, j. Further, let the the average individual's characteristics be denoted by X_A.
- Let $d(X_i, X_j)$ measure the social distance of i to j,

Empirical Model

- The decision by i to form a link with j is captured by $g_{i,j}$
- Let m_j be the total proportion of links (made or received or both) of player j
- Let v_j measure the value of player j's information aggregation
- Let X_i, X_j collect individual characteristics of players i, j. Further, let the the average individual's characteristics be denoted by X_A.
- Let $d(X_i, X_j)$ measure the social distance of i to j,
- Note that the cost of link formation must be proportional to the social distance or say $\beta d(X_i, X_j)$

Empirical Model

The the following represents the decision of player i:

$$\begin{aligned} g_{ij} &= 1(\pi_{ij}(g) > 0) \\ \pi_{ij}(g) &= \delta m_j + \alpha v_j + \beta d(X_i, X_j) + \varepsilon_{ij} \\ m_j &= \gamma d(X_j, X_A) + \eta_j \\ \varepsilon_{ij} &= \rho \eta_j + \nu_{ij} \end{aligned}$$

where δ, α, β are parameters to be estimated, ε_{ij} is the error term, ν_{ij} and η_j are independent of all the regressors and all errors are assumed to be normally distributed.

Spatial Correlation

Spatial Correlation

- The errors might still be spatially correlated.

Spatial Correlation

- The errors might still be spatially correlated.
- The error terms ν_{ij} and ν_{ik} might be correlated because they capture the error in g_{ij} and g_{ik}, both of which are i's decision to form a link with different players

Spatial Correlation

- The errors might still be spatially correlated.
- The error terms ν_{ij} and ν_{ik} might be correlated because they capture the error in g_{ij} and g_{ik}, both of which are i's decision to form a link with different players
- Similarly, g_{ji} and g_{ki} reflect j's and k's decision, respectively, to form a link with i. Since both of these decisions depend on the characteristics of i they are not independent and neither are the errors ν_{ji} and ν_{ki}.

Spatial Correlation

- The errors might still be spatially correlated.
- The error terms ν_{ij} and ν_{ik} might be correlated because they capture the error in g_{ij} and g_{ik}, both of which are i's decision to form a link with different players
- Similarly, g_{ji} and g_{ki} reflect j's and k's decision, respectively, to form a link with i. Since both of these decisions depend on the characteristics of i they are not independent and neither are the errors ν_{ji} and ν_{ki}.
- Use the corrected standard errors as suggested by Cameron, Gelbach and Miller (2011).

Data from Ghana

Data from Ghana

- Data by Udry/Goldstein over two years, fifteen modules, four village clusters in Eastern Region of Ghana.

Data from Ghana

- Data by Udry/Goldstein over two years, fifteen modules, four village clusters in Eastern Region of Ghana.
- In each village 60 couples/triples were questioned.

Data from Ghana

- Data by Udry/Goldstein over two years, fifteen modules, four village clusters in Eastern Region of Ghana.
- In each village 60 couples/triples were questioned.
- The network data was collected by asking each individual in the sample about seven randomly selected (without replacement, from the sample).

Data from Ghana

- Data by Udry/Goldstein over two years, fifteen modules, four village clusters in Eastern Region of Ghana.
- In each village 60 couples/triples were questioned.
- The network data was collected by asking each individual in the sample about seven randomly selected (without replacement, from the sample).
- Link question used is: Could you go to x if you had a problem with unhealthy crops?

Data from Ghana

- Data by Udry/Goldstein over two years, fifteen modules, four village clusters in Eastern Region of Ghana.
- In each village 60 couples/triples were questioned.
- The network data was collected by asking each individual in the sample about seven randomly selected (without replacement, from the sample).
- Link question used is: Could you go to x if you had a problem with unhealthy crops?
- The data on identity: age, religion, clan, gender, if they are the first of their family to reside in that village, experience with pineapples, wealth and soil type.

Table: Variables Measuring Presence of Link

Variable	Definition	Mean
Askprob	1 if respondent would ask match if they had a problem with unhealthy crop, 0 o.w.	0.327
Askfert	1 if respondent would go to match for advice on new fertilizer, 0 o.w.	0.303
Askplant	1 if respondent would go to match to discuss planting method, 0 o.w.	0.309
Askbuyer	1 if respondent would go to match for find a buyer, 0 o.w.	0.253
Ask	sum of the previous four variables	1.191

Table: Summary Statistics of the Respondent

age	age in years	40.079
off	1 if respondent hold an office, 0 o.w.	0.206
school_level	values from 0 to 5 for school level	1.53
pineyrs	experience in years with pineapple farming	2.125
firsthere	1 if respondent is first of his family to reside in the village, 0 o.w.	0.231
resprel	values from 1 to 16 for religion	4.272
Clan	values from 1 to 30 for clan	7.233
Gender	1 if respondent is female, 2 if male	1.454
stype	values 1 to 3 for soil type	1.887
tot_wealth	value of the nonland assets (in million cedis)	0.851

Table: Corresponding Variables for the Match

Variable	Definition
Mage	age of the match in years
Moff	value 1 if match holds an office, 0 o.w.
Mschool_level	discrete variable taking values from 0 to 5
Mpineyrs	match's experience in years with pineapple farming
Mfirsthere	1 if match is first of his family to reside in the village, 0 o.w.
Mresprel	discrete variable taking values from 1 to 16
MClan	discrete variable taking values from 1 to 13
MGender	value 1 if match is female, 2 if male
Mstype	values 1 to 3 for soil type
Mtot_wealth	value of the nonland assets (in million cedis)

Table: Variables Measuring Distance between Respondent and Match

Variable	Definition	Mean
Shhn	1 if both from the same household, 0 o.w.	0.007
Sfirsthere	1 if either both first from their families in the village, or both not the first in the village, 0 o.w.	0.655
Sresprel	1 if both have the same religion, 0 o.w.	0.272
Sgender	1 if both have the same gender, 0 o.w.	0.499
Sclan	1 if both belong to the same clan, 0 o.w.	0.304
Sptot_wealth	absolute difference in wealth if respondent is wealthier	0.448
Sntot_wealth	absolute difference in wealth if match is wealthier	0.98
Sstype	1 if both have the same soil type, 0 o.w	0.41

Table: Variables Measuring Total Links of Match

Variable	Definition	Mean
Mpin_prob	Proportion of links received by match regarding information on unhealthy crop	0.277
Mpin_fert	Proportion of links received by match regarding information new fertilizer	0.233
Mpin_plant	Proportion of links received by match regarding information on planting method	0.258
Mpin_buyer	Proportion of links received by match regarding information on finding a buyer	0.252
Mpin	Proportion of links received by match regarding any information	0.255

Mpout_prob	Proportion of links made by match regarding information on unhealthy crop	0.324
Mpout_fert	Proportion of links made by match regarding information on new fertilizer	0.304
Mpout_plant	Proportion of links made by match regarding information on planting method	0.308
Mpout_buyer	Proportion of links made by match regarding information on finding a buyer	0.252
Mpout	Proportion of links made by match regarding any information	0.297

Table: Variables Measuring Distance of Match from Average Respondent

Variable	Definition	Mean
Mdmoderesprel	1 if match has the modal religion , 0 o.w.	0.438
Mdmodeclan	1 if match belongs to the modal clan, 0 o.w	0.464
Mdpmeanage	Absolute difference in age between match and average if match is older, 0 o.w.	5.576
Mdnmeanage	Absolute difference in age between match and average if match is younger, 0 o.w.	4.908

Mdpmeantot_wealth	Absolute difference in wealth between match and average if match is poorer, 0 o.w.	0.368
Mdnmeantot_wealth	Absolute difference in wealth between match and average if match is wealthier, 0 o.w.	0.364
Mdmodestype	1 if match has the the modal soil type, 0 o.w	0.72

Table: Simple OLS results for the Variable Ask

	Ask
MPin	3.541
	(15.08)**
MPout	0.048
	-0.28
off	-0.6
	(2.68)**
Mschool_level	-0.133
	(3.02)**
pineyrs	-0.057
	(2.26)*
Shhn	1.543
	(2.87)**
Observations	790
z statistics in parentheses	
* significant at 5%; ** significant at 1%	

Table: First stage OLS results for total links received by match

	MPin
Moff	-0.006
	-0.1
Mschool_level	0.001
	-0.06
Mpineyrs	0.016
	(2.70)**
Mfirsthere	-0.097
	-1.83
Mdmoderesprel	-0.005
	-0.11
Mdpmeantot_wealth	-0.052
	-0.66
Mdnmeantot_wealth	0.004
	-0.26
Observations	133
R-squared	0.13

Table: First stage OLS results for total links made by match

	MPout
Moff	-0.088
	-1.33
Mschool_level	-0.072
	(3.10)**
Mpineyrs	0.004
	-0.65
Mfirsthere	-0.172
	(3.03)**
Mdmoderesprel	0.014
	-0.31
Mdpmeantot_wealth	0.209
	(2.49)*
Mdnmeantot_wealth	0.031
	-1.65
Observations	133
R-squared	0.23

Table: Control Function results for the Variable Ask with correct S.E.

	Ask
MPin	4.659
	(3.51)**
MPout	-1.109
	(2.00)*
off	-0.664
	(2.71)**
Moff	-0.162
	-1
school_level	-0.123
	-1.38
Mschool_level	-0.232
	(3.72)**
pineyrs	-0.07
	(2.74)**
Mpineyrs	-0.01
	-0.44

Shhn	1.774
	(3.05)**
Sclan	0.343
	(2.06)*
Mpinresid	-1.393
	-1
Mpoutresid	1.313
	(2.06)*
Observations	630

Conclusion

Conclusion

- This paper used results from game theoretic models that predict information links are more likely to be formed with players with higher levels of connectivity.

Conclusion

- This paper used results from game theoretic models that predict information links are more likely to be formed with players with higher levels of connectivity.
- The game theoretic models also suggest a way to overcome the inherent endogeneity of the previous prediction.

Conclusion

- This paper used results from game theoretic models that predict information links are more likely to be formed with players with higher levels of connectivity.
- The game theoretic models also suggest a way to overcome the inherent endogeneity of the previous prediction.
- The results from the data indicate that in the particular data set used, links are in fact formed taking into consideration both the number of links received and made by the match.

Conclusion

- This paper used results from game theoretic models that predict information links are more likely to be formed with players with higher levels of connectivity.
- The game theoretic models also suggest a way to overcome the inherent endogeneity of the previous prediction.
- The results from the data indicate that in the particular data set used, links are in fact formed taking into consideration both the number of links received and made by the match.
- The number of links made by the match is decreasing in their education and wealth level, indicating perhaps that more links are made by nodes of lower informational value. This is further reflected in the fact that the probability of forming a link is decreasing in the number of links made by the match.