Social and Information Networks

University of Toronto CSC303
Winter/Spring 2019

Week 12: April 2,4 (2019)

Announcements

Announcements

- Tyrone and I have been reading the *critical reviews* and for the most part the reviews are very good. Some students did not understand the "critical" aspect of the review and just presented a summary or abstract of the paper. Some omissions of any citation.
- We have already many reviews that we would like presented so now I am intending to us Wednesday and Friday for presentations if we can get enough volunteers. I also have to track down emails to notify students that we would like them to present.
  NOTE: Your grade is based solely on the written review so this is totally optional.
- Celebrating Steve Cook and 50 Year of NP-Completeness. Please look at information on the conference to be held at the Fields Institute. Many Turing Award winners will attend and speak. Opening lecture by Christos Papadimitiou Monday night is free. Conference costs $75 for students and is a bargain as it includes some lunches and a banquet. But registration is almost full.

Todays agenda

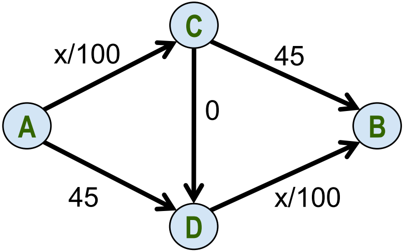
Todays agenda.

1. Finish discussion of Braess Paradox
2. Price of Anarchy and Price of Stabilty
3. Recap of course
4. Presentations of some of the critical reviews

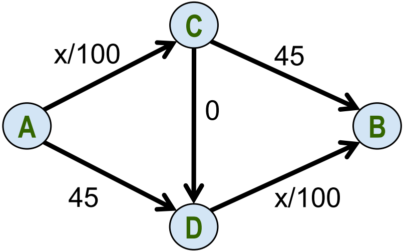
Braess' Paradox

Suppose the premier decides to build a new superhighway (or super fast rail line) and add this to the existing traffic network.

Lets even imagine that the time to traverse this new additional link is negible (and hence approximated by 0 time). It seems that this can only improve the life of commuters. So lets add a directed link from C to D in our example traffic network.



Braess' paradox continued



Claim: There is a new unque NE. Everyone now will want to take the route $A \to C \to D \to B$. And the individual commute time of this NE is 80 minutes! That is, by building the new superhighway (rail link) everyone has an additional 15 minutes of commuting.

Proof of claim for Braess' paradox

- Everyone taking $A \rightarrow C \rightarrow D \rightarrow B$ is an NE. This can be seen by considering any individual wanting to deviate. Deviating by taking the direct (A, D) edge is worse (for the one person deviating) than taking the indirect path to D via C. So the potential deviating commuter will want to first go to C and then from C, it is better to take the indirect path (via D) to B than taking the direct (C, B) link.

Another equivalent way to state this paradox is that in some traffic networks, closing a road or rail link might speed up the commute time! And this has been observed in some cases. Of course, all this assumes that individuals will find their way to an equilibrium.

The new link and social welfare

Is there any sense in which this new link can be beneficial? Consider the social welfare that is now possible with the new link. Note that we now have three paths amongst which to distribute the load.

Claim: The following is a socially optimal solution:

- 1750 take $A \rightarrow C \rightarrow B$ route
- 500 take $A \rightarrow C \rightarrow D \rightarrow B$ route
- 1750 take $A \rightarrow D \rightarrow B$ route

Society wins but some people lose

What is the social welfare of this solution? We have

- 500 commuters taking the $A \to C \to D \to B$ route will each have travel time 45 minutes saving 20 minues each in comparison to the 65 minute commute time without the new 0 cost link.
- On the othert hand, the $1750 + 1750 = 3500$ commuters taking the more direct $A \to C \to B$ or the $A \to D \to B$ routes will each have travel time 67.50 minutes incurring an additional 2.5 minutes of commute time.

So the *total time* saved is $(500 \times 20 - 3500 \times 2.5) = 1250$ minutes each way, each day. **On average** (over the 4000 commuters), it is a saving of $1250/4000 = .3125$ minutes per commuter. If this doesn't sound sufficiently impressive, suppose time was being measure in hours; that is, we can scale the edge costs by any fixed factor.

So do we build the new road or railway link?

Even if the cost of the new link is not a factor, do we build the new link?

So do we build the new road or railway link?

Even if the cost of the new link is not a factor, do we build the new link?

Some of the commuters now have incurred some additional travel time and will explore other routes. We view this as an unspecified random process with different individuals exploring new routes from time to time. Will they eventually return to the solution without the new link where everyones commute time was 65 minutes, or (as game theory suggests) will they (by self-interest) eventually converge to the unique Nash Equilibirum (NE) where everyone takes the $A - C - D - B$ route?

The unequal partition into the three routes $A - C - B, A - D - B$ and $A - C - D - B$ is not an equilbrium but it is a social optimum in this expanded network whereas the equal partition into the two $A - C - B$ and $A - D - B$ routes was a socially optimum NE in the network without the $C - D$ road.

Understanding the partition into 3 routes

How do we argue the previous solution is a social optimum and how do we find this partition of routes?

There is something very symmetrical about the network that the new link can now exploit. Note that we can we equalize the total time used between going from A to C and from going from A to D (either directly or via C) by having 2250 going to C (with 1750 going on directly to B and 500 taking the $C - D$ road) and 1750 going to B via the $A - D - B$ route.

This can be determined by solving a quadtratic equation to determine the x commuters who will initially go to C and the $4000 - x$ that will initially go to D. By the network symmetry and by redistributiing the load via the $C - D$ road, this becomes the same of the for $(4000 - x)$ commuters to take the $A - C - B$ route.

Total time is: $x \cdot \frac{x}{100} + (4000 - x) \cdot 45 = .01x^2 - 45x + 180000$.
Taking the derivative and setting it to 0, we get: $.02x - 45 = 0$ resulting in the desired solution that $x = 2250$. That is, $4000 - x$ will take the $A - C - B$ route, 1750 will take the $A - D - B$ route and that means redirecting 500 from C to D.

How could the government obtain the socially optimum solution?

If the government wants to achieve this social optimum (say to show they didn't waste money building this super fast road), they will somehow have to "encourage" or impose this solution in some way since the socially optimum solution is not an equilibrium. (What individual commuter would want to take the $A - D$ or $C - B$ roads if everyone else was taking the $A - C - D - B$ route?) There are different ways to encourage or impose a solution. For example:

- One implicit way to hopefully influence drivers to converge towards the socially better equilibrium is to place a toll on the new link; by adjusting the pricing on the new link, the idea would be that commuters who have the money and value their time more would start taking the new route. (Or similarly, they could sell passes to the new road, similar to selling passes to the HOV lanes.)
- They could limit the number of commuters taking the $A - C$ road

The Tragedy of the Commons and the Price of Anarchy

If we believe commuters will converge to a NE, then allowing commuters to act in their own interest has a "price" (with respect to social optimality). In this network road example, the price is the additional total time (1250 minutes) to commute.

This price of self interest in this or any setting where self interest is a factor is often refered to as the Tragedy of the Commons.

In the computer science literature (algorithmic game theory), there is a quantitative measure of the price we pay for self-interest with respect to social optimality. In general, there can be many pure and mixed NE.

The Price of Anarchy (POA) for any such specific "game" (where the social objective is a cost function) is a worst case ratio measuring the cost of stability; namely, taking the worst case over all NE solutions S, it is defined as : $\frac{cost(S)}{cost(OPT)}$ where OPT is an optimum solution.

The Price of Anarchy continued

The Price of Anarchy was introduced by Papadimitriou.
Aside Christos Papadimitriou is giving a public lecture, Monday night (May 6) as the initial event in the symposium celebrating Steve Cook and 50 years of NP-completeness (May 6-May 9).

For a more optimistic perspective there is also a Price of Stability defined as: $\frac{cost(S)}{cost(OPT)}$ where now S is a NE solution having the least cost.

Returning to the specific setting of network congestion, the following two results (due to Roughgarden and Tardos) are early seminal results in algorithmic game theory. For *all congestion networks with linear cost functions*:

1. The POA is no more than $\frac{4}{3}$
2. By adding a new link, the change in the social optimum cannot increase by more than a factor of two.

New topic: Kidney exchanges

Although this will not be on the final exam, the topic of kidney exchanges is technically interesting and , of course, critically important for many people.
Some facts:

- In the US, each year there are 50,000 new cases of potentially lethal kidney disease.
- There are two possible treatments: dialysis or transplant.
- Transplants can come from live donations or from transplants for someone who has just died (e.g., in car accident). All else being equal, live donations are much mor successful.
- Each year there are $\approx 10,000$ transplants from someone deceased and ≈ 6500 from live donations.
- The waiting list for a transplant in the US is $\approx 75,000$ people who usually wait between 2 and 5 years. During this waiting time, ≈ 4000 people die each year.

More facts concerning kidney exchanges

Live donations are possible since everyone has two kidneys and only one is needed. Moreover, when people incur kidney diseasse, usually both kidneys are effected so the “additional kidney” is rarely needed.

However, people are reluctant to donate kidneys and live donatuions usually come from close relatives and friends.

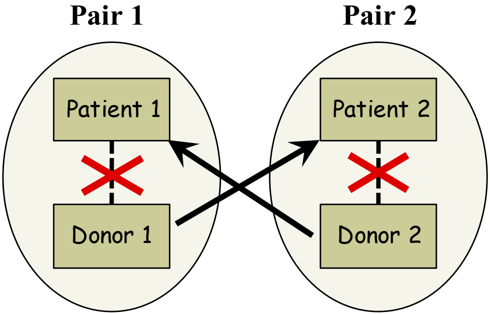
There are many biological compatability requirements in order to do a transplant so there is often no one availble and willing to do a donation.

- Blood compatability
- Tissue compatability

Even if possible, some donator-recipient transplants are better than others.

Pairing up transplants

So if a willing donor for a recipient is not compatible (or if the match is not that great), there may be another reciptient-donor pair that are having the same issue and are willing to do a 'swap". Consider the following possibiliy for a pair swapping:

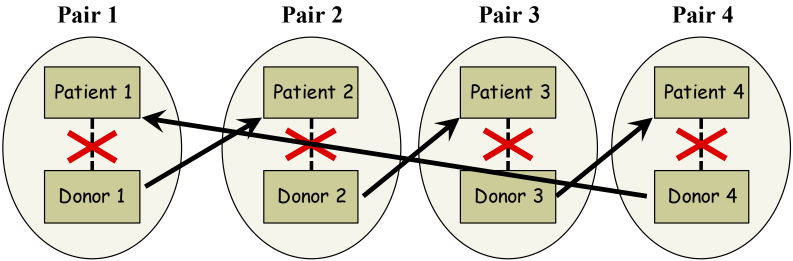


Here an edge means that the Patient (i.e. the recipient) and Donor are compatible. Edges can be weighted to reflect some objective as to how good is the match. The weight could also reflect geographic distance.

Extending to bigger cycles

The idea of pairs swapping as just illustrated was first proposed in 1986 and only realized in 2003.

This idea has been extended to bigger cycles as in the next illustration:



How practical are such swaps and cycles?

The are "logistical" issues that impact the practiicality of such swaps and cycles, and the bigger the cycle the more problematic logistically.

What is a potential donor, say Donor i renegs (or dies, or gets ill) once his/her paired recipient Patient i has already received their (from Donor $i-1$) kidney from the person with whom they are compatable? Now Patient $i+1$ has lost ia valuable resource his/her (the intended Donor they brought to the exchange) if Donor $i+1$ has already given their kidney to Patient $i+2$.

This requires that the donation and transplant must all basically be done *simultaneously*. For cycles of length k, this requires $2k$ simiultaneuous operations, where each translantation reqyuires both a donation and transplant operation.

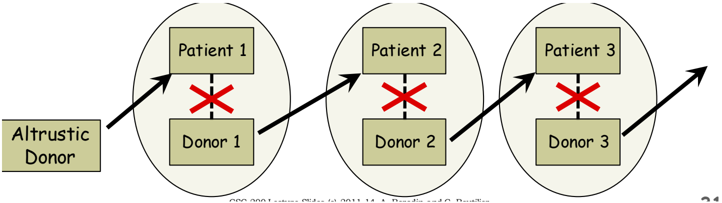
Furthermonre, live kidneys from donors travel best inside the donor, so need these operations to be geogephically close (i.e. same or nearby hospital). Note: Some hospitals will not accept organs transplanted by air.

The net effect is that this severly limits the length of cycles that cab be

Altruistically initiated donor chains

Suppose we have one altruistic donor who is willing to donate a kidney without having someone with whom he/she wishes to be paired? Once there is such an altruisic donor, we can eliminate the need for simultaneity.

After we have an altruistic donor, we can proceed in what potentially can be an arbitarily long chain as below. Here each Patient must still be willing to bring a willing Donor to the exchange. But now if some donor renegs, etc, the next reciptient has not lost their paired donor.



There has been at least one chain of length 30 (ending in February 2012) and some chains may be still be ongoing.

Some final comments

Given all the biological and logistical (and incentive) issues the area of kidney exchanges is an area that requires efficient algorithmic solutions..

We are talking about pretty large scale networks; i.e., say tens of thousands of nodes when considered nationwide.

When restricted to pairs, this is a (possibly weighted) matching problem in a non-bipartite graph. When we introduce cycles and chains the problem becomes much harder. This becomes a matter of computing "practically feasible" cycles and chains.

In addition, the market is not a static network. There are arrivals and departures. This raises other issues:

- Is it better to use a current match, or wait for new donors and recipients to arrive?
- When an altruistic donor arrives, do you use up that valuable resource now or wait for a better match that might lead to a longer chain.
- Are there incentive issues for say hospitals to want to do more of the transplants by themselves than join in a borader exchange?

End of Monday, April 2 Lecture

Announcements:

- Tyrone and I will be checking that all grades have been properly recorded. Please keep emailing us (at the email for the course) if any grade is missing and/or not properly recorded.
- The online course evaluation is now open. I encourage you to do an evaluation.

  Students can complete their course evaluations one of two ways: 1) Via the personalized e-mail link that is sent to them automatically by the online course evaluation system at the start of the student evaluation period. 2) Within the Course Evals tab in Quercus Both options are available to students at any time during the student evaluation period.
- I will be available often for office hours during exam period but I am not available April 17-23 inclusive. Please email me to be sure of a meeting time or take your chances and just drop in.

Today (and Friday) we will have some presentations. I will discuss the final exam on Friday. First a brief recap of the course.

A recap of the course

We can keep this recap in mind to see how the presentations fit into the theme and topics in this course.

I would say that the central theme of the course is the attempt to more precisely model sociological phenomena. This includes the relatively less studied (in the course) "information networks" (e.g., the web) as it is humans that create this network. The way we link and rank documents, and "navigate" within this network of documents fits into social networks.
Aside: I am now looking at a relatively new paper as to how power laws emerge in the graph of routers and other aspects of the internet.

The main mathematical framework (and hence the course name) centers around networks. Modeling social networks presents significant challenges and in many cases, there are only initial insights and we are far from realistic models and analysis of social phenomena.

Recap continued

To the extent that current social networks are often extremely large, it is necessary to be able to "think algorithmically" while appreciating the fundamental insights and studies that have evolved and continue to evolve from sociology, economics, biology, physics, and other fields. Being able to reason about stochastic models is also obviously necessary.

As the text often emphasizes, in what may be called algorithmic social networks, the approach taken follows what we see in other sciences. Informed by real world networks and phenomena, we formulate precise models, draw some insights and possibly some preliminary conclusions, and then calibrate the model and insights against real world or synthetic data. Based on the experimental results, we are then able to iterate the process; that is, modify the model and continue to draw insights and again evaluate by experiments.

Recap continued

The text properly cautions that these models are just that, *only models* of real world network behaviour and that we are often far from having confidence in any preliminary conclusions.

In some cases, it is suprising how much information one can obtain just from basic network models and assumptions. A good example is the identification of romantic ties in the Backstrom and Kleinberg paper and the labeling of strong and weak ties in the Sintos and Tsaparas paper. But, of course, the more we know about the content relating to the nodes and edges in a network, the more we should be able to make informative findings.

Some of the major topics in the text and the course

Here are some of the major topics in course:

- The concept of strong and weak ties and their relative role in obtaining "social capital".
- Different types of closing of triangles: triadic closure, focal closure, membership closure.
- Homophily and influence. To what extent are our frienships derived from similar interests and behaviour vs that our friendships are influencing our interests and behvaviour. This is a central issue in social relations and one where any findings can be controversial. For example, recall the issue of whether or not "obesity is contagious" to some extent.
- A number of topics relate to different equilibrium concepts. We discussed structural balance, Schelling's segregation model, self-fulfilling expectations with regard to direct benefit effects, balanced outcomes in bargaining networks, stable matchings, and Nash equilibria in a congestion network. This also relates to collective action as in direct network effects,

Some major topics continued

- A number of topics relate to navigation in a social network and in particular to the small world phenomena based on goegraphic or social distance. This also was related to power law distributions in social and information networks.
- Influence spread in social networks and disease spread in contact networks. Cascades.
- Am I missing any major themes that we discussed?

What I am hoping to develop better next time we present this course

The text is now several years old but still an excellent text. We presented a few topics outside of the text material, namely the problem of influence maximization, and stable matchings.

While most topics related explicitly to networks, there were a few topics for which I would like to better understand the role of underlying networks. The chapters on cascades and direct benefits do not explicitly relate to networks except in some very elementary way. Namely, in the cascade discussion, we can think of the network as a line network where node i has a link to all nodes $j < i$. In the direct benefits model, we can assume a complete network so that we are aware (in some anonymous and summarized way) about all other buyers.

Most of the emphasis in the course was on static networks whereas real world networks are very dynamic.

It would be good to understand better viral spread in online networks and how much they influence (for example) the political process.