5. Social Welfare and Bittorrent credit dynamics

Modelling Social Interaction in Information systems

http://davidhales.com/msiis

David Hales, University of Szeged

dave@davidhales.com

Recap from previous lectures

- The Prisoner's Dilemma showed how a contradiction between individual utility and collective utility could be represented
- We saw how a repeated game could lead to high collective utility through the tit-for-tat (TFT) strategy
- We looked at the bittorrent (BT) protocol to examine how TFT was implemented
- We noted that BT relied on seeding which was not part of the TFT approach and had no incentives associated with it

Recap from previous lectures

- In the BT paper by Cohen (2003) he argued the protocol attempted to attain Pareto efficient (PE) outcomes
- PE is one way to characterise a “collectively good outcome”
- In general for any socio-economic system we can construct any function and define it as “collectively good” depending on what we want
- Sometimes called a “Social Welfare Function” (SWF)
- **Warning:** I will talk here about SWF’s in a very informal and simplified way. Different areas of “welfare economics” use various assumptions and definitions < you have been warned!

Pareto efficiency

- As we have seen a given outcome is *not* Pareto efficient if:
  – There exists some other arrangement such that at least one agent can improve its utility
  – Without any other agent having to reduce its utility
- Any change in arrangements that improves utility in this way is called a "Pareto Improvement"
- If any change of arrangements makes anyone worse off then it is not PE by definition
- In PD example everything is PE other than mutual defection
- PE is interesting because it shows how a system that is not PE might move towards PE through actions compatible with individual agent preferences and behaviours
- PE says nothing about equality or total utility and in general systems contain many (possibly infinite) PE states
- Formulated by Vilfredo Pareto (1884-1923)

Social Welfare Functions

- A SWF (as we use it here) may or may not conflict with individual agents' preferences
- For example, we saw that in the PD if we defined the SWF as the sum of agent utilities
  - Rational utility maximisers would defect
  - Producing an equilibrium far from max SWF
- However, if we defined the SWF as maximum equality between agents then mutual defection *would* maximise the SWF (as well as mutual cooperation)

Social Welfare Functions

- There are many SWF proposed by economists. Here are three simple ones:
  - Sum of all utilities (utilitarianism)
  - Max-Min (maximise worst off agent)
  - Max*equality (maximise product of utility and equality)
- A SWF itself does not tell us how agents might behave such that they would maximise it
- But much work related to SWF considers how this can come about

Utilitarianism

- Sum all the utilities of all the agents
- Intuitive and simple to understand (founder of utilitarianism: Jeremy Bentham 1748-1832)
- In the PD example this SWF would be maximal with mutual cooperation
- However, it just happens that in the PD game this leads to perfect equality
- But this might not be the case in other scenarios
- For example, imagine a situation in which one agent could have a utility of X with all other agents having zero. If X was > than the sum of all agent utilities in all other arrangements then it would maximise utility

Max-Min

- Maximises the utility of the lowest utility agent in the population
- Inspired by ideas of John Rawls (1921-2002)
- Does not maximise total utility or equality
- In the PD mutual cooperation would result from Max-Min
- Of course this happens to maximise the total utility too but this may not be the case in other scenarios
- Consider a situation in which the poorest agent's utility was maximised at the expense of all the other agents utilities – let's all take pay cut of 50% to increase the poorest agents pay by 1%?

Book: Rawls, John (1971). A Theory of Justice. Cambridge, Mass.: Harvard University Press.

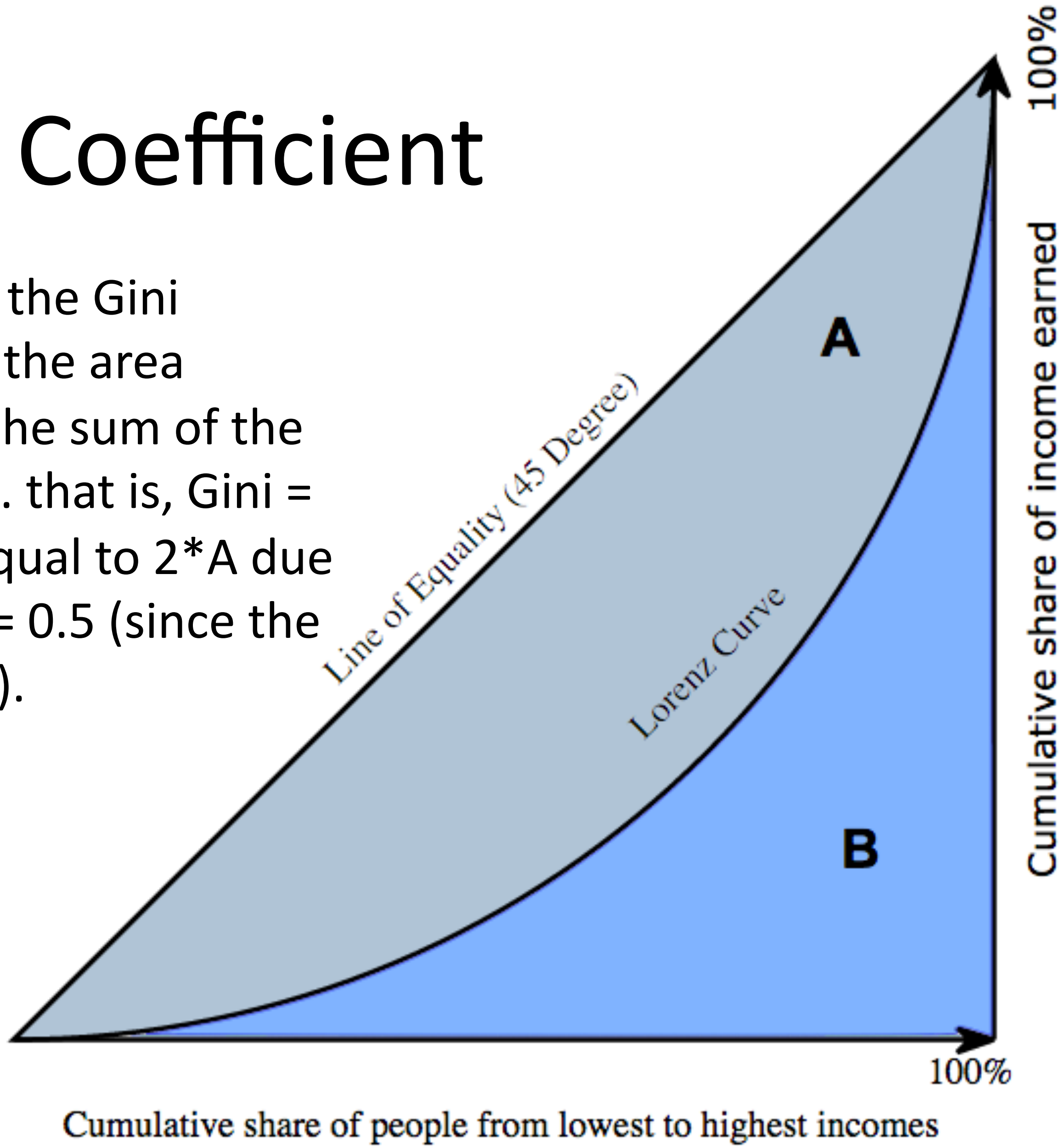
Max*equality

- Maximise the total utility multiplied by a measure of equality over all agent utilities
- SWF introduced by Amartya Sen (1933-):
  - Average income * (1 – G)
  - Where G = the Gini coefficient for inequality
  - G goes from 0 (total equality) to 1 (total ineq.)
- This would select mutual cooperation in the PD example
- However, in scenarios where more inequality would lead to *much* higher utilities on average then higher levels of inequality would be selected

Book: Amartya K. Sen, 1970 [1984], Collective Choice and Social Welfare, North Holland; Third printing edition.

Aside: Gini Coefficient

The graph shows that the Gini coefficient is equal to the area marked A divided by the sum of the areas marked A and B. that is, Gini = A / (A + B). It is also equal to 2*A due to the fact that A + B = 0.5 (since the axes scale from 0 to 1).



From: http://en.wikipedia.org/wiki/Gini_coefficient

Social Welfare Functions

- I have given a very crude overview of some SWF that emerge from economics
- The way these relate to individual agent choices is an area called "social choice theory"
- This is a huge area of incredible sophistication and debate
- Characterising various economic, social, political and philosophical arguments and approaches
- In general a lot of the work assumes agents are "rational" in an economic sense of trying to maximise their individual utility – but not all work!
- Designing rules that make rational individuals select some social optimum is termed "mechanism design" in general and "computational mechanism design" in computers

Social Welfare Functions

- The SWF we looked at all rely on some way to:
  - characterise agent utility
  - make comparisons between utility
  - argue that one social outcome is better than another
- They can be used to justify government policies by arguing that a particular policy will increase social welfare
- But can also be used as design objectives in distributed systems – remember Cohen (2003) talking "Pareto"
- If we design a distributed system we are probably working with a "kind of" SWF even if we are not aware of what it is
- That is, we generally have some idea of what a good outcome would be over a bad outcome – although perhaps not always!

Aside: Rawls' Theory of Justice

- John Rawls is famous for his modern take on "social contract theory" (remember Hobbes?)
- assume we wish to specify the kind of society that is just and good
- but we stand outside the society and don't know what role we ourselves would play
- we are ignorant of what endowments, knowledge, capacities and position we would hold ("veil of ignorance")
- what rules / norms would we accept as just and fair? i.e. what would we accept as "collective good"?
- The idea is that given the "veil of ignorance" agents would rationally chose a society (it's rules, laws etc) that is "Just and Fair"
- There is an argument that it would be rational to maximise the least well off

Credit Dynamics in a BT private tracker Community

- Look at the phenomena of credit squeezes in a private BT community
- Use a simple “toy” (agent-based) model to capture credit dynamics
- Do some measurements of an actual private community
- Look at inequality and throughput of the system as a whole

Hales, D., Rahman, R., Zhang, B., Meulpolder M., and Pouwelse, J. (2009) BitTorrent or BitCrunch: Evidence of a credit squeeze in BitTorrent? Proc. of the 5th Collaborative Peer-to-Peer Systems (COPS) Workshop

Public Trackers

- BitTorrent uses Trackers to index swarms
- Public trackers let anyone join a swarm
- Sharing is incentivised via a form of tit-for-tat
- However there is no incentive for:
  - Seeding (uploading after file is downloaded)
  - Capping (creating and injecting a new file)

Private Trackers

- Private Trackers have emerged more recently
- Only allow registered users to join swarms
- Track upload / download of each user
- Keep centralised accounts for each user
- When users download much more than upload they may be kicked out
- Different schemes: ratio, credits, points etc

Private Trackers - Credit

- Consider a scheme based on credits
  - Uploading 1MB earns one credit
  - Downloading 1MB costs one credit
  - A user with no credits can't download
- Users must be given some initial credit
- In fixed size pop. total credit remains constant
- Similar to a fixed supply of money in an economy (loose analogy!)

Private Trackers - Credit

- How much credit should be put into the system?
- How would it effect the efficiency of the system?
- When do credit squeezes occur?
- How can they be avoided?

We define a credit squeeze as a situation in which, due to lack of credit, the efficiency (throughput) of the system is significantly reduced.

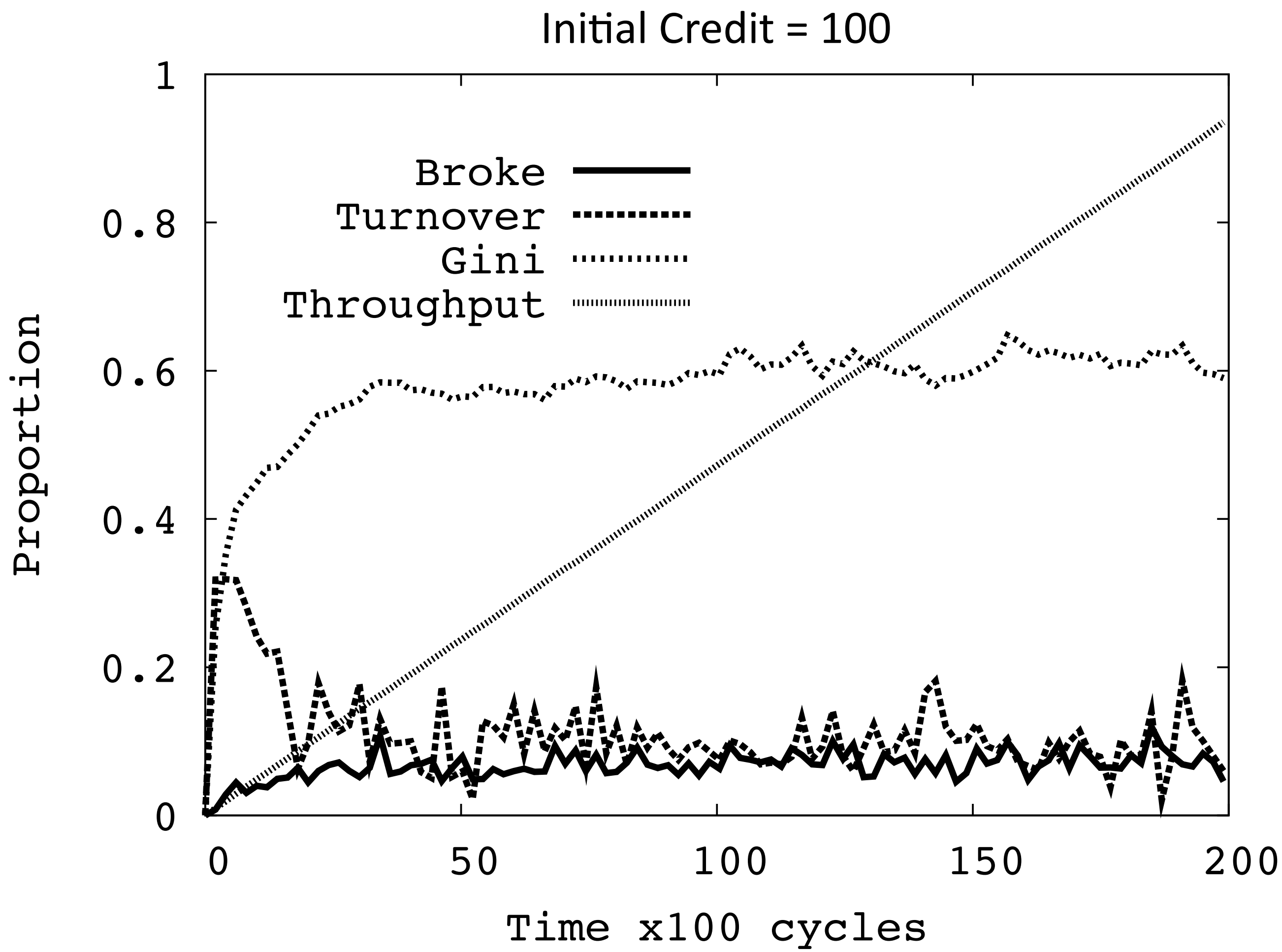
BitCrunch Model

- Highly abstract and simplified agent-based model
  - All nodes have equal upload / download
  - Equally interested in all swarms on the tracker
  - Always uploading to one swarm (seeding)
  - Always downloading from one other swarm (leeching)
  - No modelling of tit-for-tat or free-riding
  - Always online, fixed population size
  - If run out of credit (broke) must wait until earns some via upload before being allowed to download
  - Swarms assumed to share upload "perfectly"

BitCrunch Model – baseline runs

- Parameters:
  - 500 peers, 100 swarms
  - Peer upload and download capacity = 1 unit
  - Each file shared in each swarm = 10 units size
  - One simulation cycle = each swarm processes one unit of time
  - Run for 20,000 cycles (x10 runs)
  - For initial credit per peer of 1, 10 and 100 units

Typical baseline simulation run



Baseline simulation results

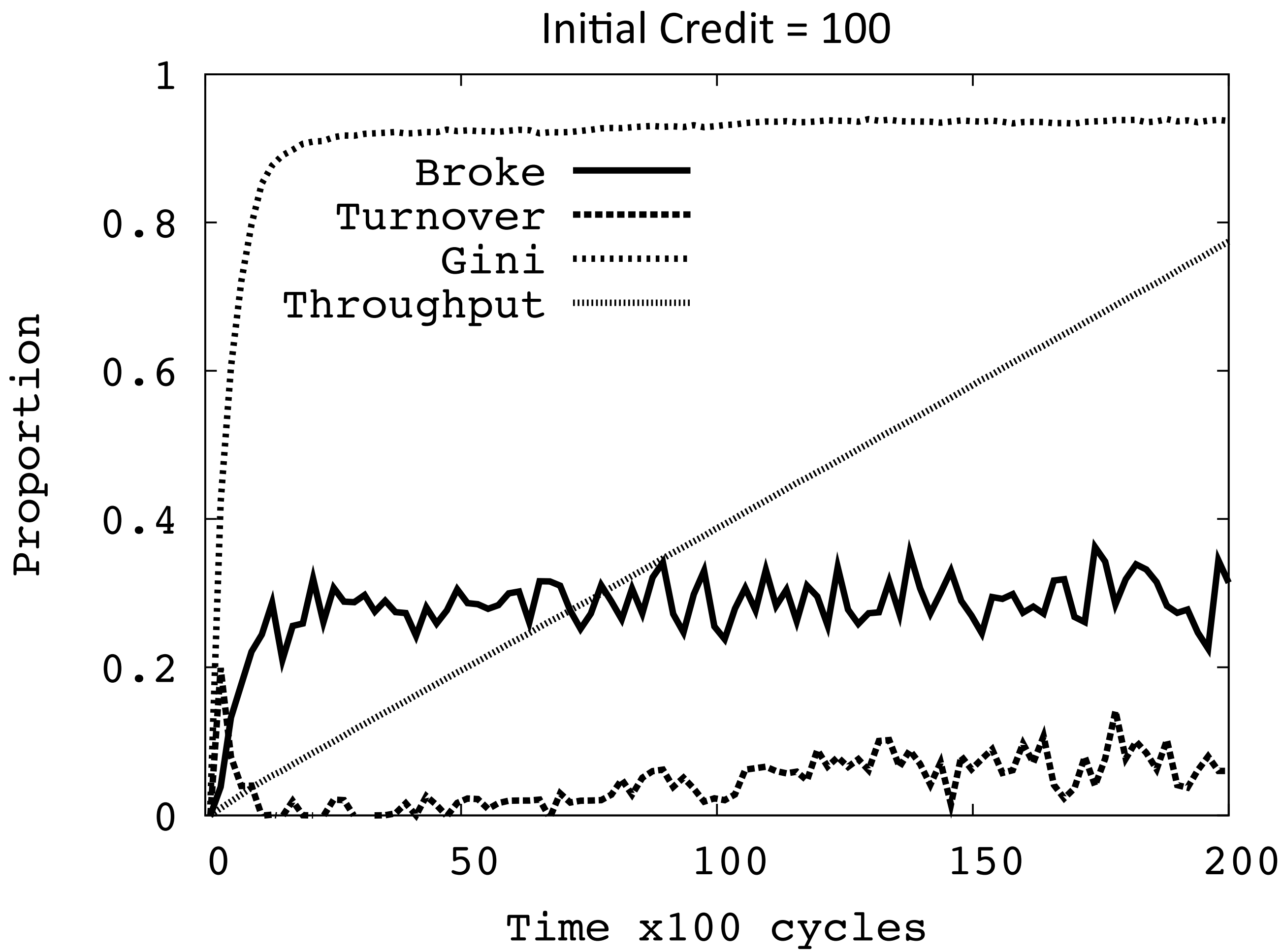
C	T	β	G	φ
1	0.58	0.36	0.87	0.84
10	0.81	0.20	0.77	0.43
100	0.97	0.06	0.59	0.10

- C = initial credit per peer
- T = total throughput = total number of units uploaded, as proportion of maximum possible (infinite credit), by end of run
- β = proportion of nodes that are "broke" (zero credit) ave. per cycle
- G = Gini measure (simple measure of inequality of credit where 0=complete equality, 1=complete inequality) average per cycle
- Phi = turnover of top 10% of peers ranked by credit (credit mobility)

Unequal capacities runs

- To determine what happens when some nodes have different upload capacities
- Parameters (same as baseline runs but):
  - All peers download capacity = 10 units
  - 10% of peers upload capacity = 10 units
  - 90% of peers upload capacity = 1 unit
  - Examined a (1.5 credit) seeding bonus approach to dynamically introduce more credit into the system

Typical unequal capacities run



Unequal capacities simulation results

C	T	β	G	φ
1	0.56	0.39	0.90	0.82
10	0.71	0.32	0.93	0.44
100	0.77	0.29	0.94	0.06
100++	0.97	0.01	0.71	0.00

- C = initial credit
- T = total throughput = total number of units uploaded as proportion of maximum possible (infinite credit)
- B = proportion of nodes that are "broke" (zero credit)
- G = Gini measure (simple measure of inequality of credit)
- Phi = turnover of top 10% of peers ranked by credit (credit mobility)
- 100++ indicates initial credit of 100 with 1.5 credit seeding bonus

Observations

– Even in a trivial model where all peers have the same capacities and altruistic behaviour, all swarms have equal popularity and all peers start with equal credits, the performance of the system may be inhibited by credit shortages
– Adding extra capacity to the system, in the form of upload and download, can actually reduce the performance by creating credit shortages
– By injecting new credit into the system in the form of a ``seeding bonus'' a credit squeeze can be ameliorated when peer capacities are unbalanced.

Statistics from a Private Tracker

Day	T	Δ	Δ_0	δ	S/L
1	48	24	17	0.23	26
2	40	20	15	0.25	26
3	50	25	12	0.16	25
4	67	33.5	17	0.17	25
5	52	26	19	0.24	25
6	46	23	15	0.21	25
7	87	43.5	17	0.13	25
Ave.	56	28	16	0.19	25

Approx. 50,000 peers per day, 10,000 swarms,
access to credit balances of top 10%. Tracker uses 1.5 seeding bonus.

T = throughput in TB over all swarms
Δ = total credit increase that day in the entire system
Δ_0 = total credit increase for top 10% of peers
δ = minimum fraction of credit increase that goes to top 10% of peers
S/L = seeder to leecher ratio over all swarms
Turnover of top 10% (not shown) is always low at less than 0.2%.

Statistics from a Private Tracker

- Indicates "rich getting richer" since top 10% are getting a lot of the new credit
- High Seeder / Leecher ratio suggestive that a credit squeeze is happening for many
- But need more information to verify this
- Would be interesting to see what happened to throughput if there was a "free leech day" or seeding bonus was increased

Summary

- Private trackers using "ratio enforcement" policies appear to be ad-hoc and various
- But can have dramatic effects on efficiency
- Too little creates credit squeezes = lower efficiency (a "crunch")
- But too much credit could encourage free-riding = low seeding (a "crash") – not modelled here (see paper in readings)

Take home message

- We can specify many possible social welfare functions
- Several versions have been explored in economics and beyond
- We can explore how credit policies effect the performance of a BT community using an ABM
- If we design a system we might want to consider what our social welfare function could be

People: all major figures in the history of ideas

- Jeremy Bentham (1784-1832), British philosopher, jurist (law), social reformer
- Vilfredo Pareto (1948-1923), Italian engineer, sociologist, economist
- John Rawls (1921-2002), American moral & political philosopher
- Amartya Sen (1933-), India economist, philosopher (Nobel prize economics)

Readings and Questions

- Papers:
  - Rahman, R. and Hales, D., Vinko, T., Pouwelse, J. and Sips, H. (2010). **No more crash or crunch: sustainable credit dynamics in a P2P community**. International Conference on High Performance Computing & Simulation (HPCS 2010)
  - Ian A. Kash, John K. Lai, Haoqi Zhang, and Aviv Zohar. (2012). **Economics of BitTorrent communities**. In Proceedings of the 21st international conference on World Wide Web (WWW '12). ACM, New York, NY, USA
- Questions:
  - Do you think the results presented here are of any practical value for creating BT community rules?
  - Rather than throughput what else could we measure to represent performance of the system
  - What would the rules of a BT private community look like if you used Rawls' theory of Justice?
  - Can you think of another simple SWF and how it would relate to the PD game and other SWF? Does it select Pareto efficient outcomes?