The impact of negative links: Theory and evidence

Xiannong Zhang[*]

April 2021

Extended Abstract

Preliminary draft

Please do not circulate

In much of the economic literature of networks, links are typically given positive meanings. However, in many applications, annotations of links are needed to understand the full picture. Links with negative meanings are crucial in modeling conflicts with alliances, bullying with defending, and trust with distrust. This paper studies theory and behavior related to networks with both positive and negative relations.

I start with a simple game-theoretical model to demonstrate the properties of stable network structures. Agents in this model decide which sign to assign to each of their links; a positive sign requires consent from both agents sharing that link and is costly, while a negative sign is the status quo. The key feature of this game is that agents with more positive links (allies) have opportunities to exploit agents with fewer positive links through negative links (conflicts). This feature results in an arms race in positive links where agents try to get an advantageous position when confronting enemies.

The implication of this result is discussed in two applications. In the alliance network application, I assume positive links do not bring any direct utility but rather serve as methods to increase one's trophies through negative links. I first show the existence of pairwise stable networks and then characterize stable network structures for a general set of payoff functions. Pairwise stable networks are assortative and have concentrated degree distribution. Strong stable networks always include exactly one agent who is an enemy to all other agents, and all other agents are allies. In the heterogenous-agent model, I found that segregation only happens among agents with many allies in pairwise stable networks. On the other hand, strong stable networks can only be one of the two extreme cases, where agents are either completely segregated by type or belong to a dense assortative component.

The second application discusses friendship-bullying networks in classrooms, where the assumption that positive links do not bring direct utility is relaxed. I vary the relative gain of bullying activities and compare stable network structures with different bullying norms: when bullying is more tolerated, friendship networks share the same properties as alliance networks. Children feared being left alone, so they are forced to join some peer groups. On the other hand, when bullying is suppressed,

[*]Department of Economics, Washington university in St. Louis

friendship networks are more star-like. Some children share friendships with many others, and some are relatively isolated.

Beyond theory, empirical studies on negative links face difficulties. Network formation is hard to study in the field due to confounding variable problems; trying to identify the effect of negative links is even more challenging. Therefore, I designed a battery of experiments to identify the impact of negative links in a well-controlled environment. In these experiments, each subject play two network formation games. In the first game, agents receive points only through their positive links. The second game is designed as close to the first game as possible, except that players may win or lose points through negative links. The goal of this experiment is to quantify the behavioral changes when negative links are introduced, with the hope of providing an explanation for the formation of real-world signed networks.

To focus on the differences in network formation associated with negative links, I did not introduce any economic interactions among agents once the network is formed. Contrary to experiments where agents interact on a fixed network, this is a “pure” network formation game where network structure is the only payoff-related variable. It is natural to ask how do subjects access their benefits derived from their network positions in this context. Do they react more to immediate benefit changes (myopia), or do they take possible subsequent changes into account (farsightedness)? Data at the individual level points in a clear direction: compared to the first game where farsighted motives dominate, subjects become much more myopic once the negative links are introduced. While the proportion of farsighted actions remained unchanged, the proportion of myopic action significantly increased by 20%. Data at the aggregate level reaches a similar conclusion. The duration of farsighted stable structures drops astonishingly from 87% to 37%. Meanwhile, the duration of pairwise (but not farsighted) stable structures significantly increased from 1% to 24%. I further discussed the implication of this result and how it affects our understanding of the formation of signed networks.