

Negative Links

Xiannong Zhang*

September 2021

Extended Abstract

In the economic literature on networks, links are typically associated with payoffs benefits. However, in many applications such as international relations or school bullying, links could yield negative consequences. This paper studies the relationship between positive and negative links, in both a theoretical framework and in an experimental setting.

I start with a theoretical network-formation model to describe the properties of stable network structures. All agents are linked to each other, but they make decisions on which sign to assign to each of their links: a positive link, indicating an alliance or a friendship, requires mutual consent and is costly, while a negative link is assumed to be costless. Any pair of agents that are not connected by a positive link are assumed to be connected through a negative link.

I explore the implications of this model in two applications. In the *alliance network application*, I assume agents are engaged in conflict with enemies, and the payoff generated by these conflicts is increasing in their own degrees and decreasing in the enemy's degree. In this setting, I first show the existence of pairwise-stable networks, and then I characterize stable network structures for a general set of payoff functions. Pairwise-stable networks are assortative, meaning agents with similar degrees tend to share positive links. Moreover, most agents in a pairwise-stable network have the same degree, while the remaining agents have a lower degree. Strong-stable networks always include exactly one agent who is an enemy to all other agents, while all other agents are allies of each other. In a heterogeneous-agent extension of the alliance network application, I show that homophily only arises among high-degree agents, while low-degree agents become allies of each other regardless of their type.

In the *school-bullying application*, I assume agents engage in two types of activities that directly affect their utilities: friendship and bullying. I characterize stable network structures as the relative gain from bullying varies: when bullying is more tolerated (the relative gains are greater), friendship networks share the same properties as alliance networks. On the other hand, when bullying is suppressed (the relative gains are lower), a star-like structure arises in stable networks. Some agents become friends with many others, and some remain relatively isolated.

Empirical studies on network formation face well-known practical difficulties due to confounding factors present in the field. Therefore, I design a battery of experiments aimed at identifying the impact of negative links in a well-controlled environment. In these experiments, each subject plays two "pure" network formation games. In the first game, agents receive points only through their positive links. In

*Department of Economics, Washington University in St. Louis

the second game, players may win or lose points through negative links. The goal of this experiment is to identify and quantify the behavioral changes caused by the introduction of negative links.

In such a network-formation setting, I observe whether agents respond more to immediate benefits (myopia), or they take potential future benefits into account (farsightedness). The results indicate that, compared to the first game where farsighted motives dominate, subjects become substantially more myopic once negative links are introduced. While the proportion of farsighted actions remained unchanged, the proportion of myopic action increases by 20%. The aggregate-level data point to a similar conclusion. The duration of farsighted stable structures (defined as the percentage of game time in which subjects maintain a farsighted stable structure) drops from 87% to 37%. Meanwhile, the duration of pairwise (but not farsighted) stable structures increases from 1% to 24%. I discuss the implication of these results and how they affect our understanding of signed network formation.