Network Formation Games

Game Theory · Module 8 — Network Games

Node 8.5

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Network Games **Node:** 8.5

In Nodes 8.1–8.4, the graph G was exogenous: given to us as a fixed interaction structure, with analysis focused on how equilibrium behavior depends on that structure. This node flips the question: **what if the graph itself is a strategic variable?** Players choose which links to form, and the resulting graph determines both the interaction structure and the payoffs. The foundational model is Jackson and Wolinsky (1996), which introduces **pairwise stability** — a bilateral equilibrium concept for network formation — and the central result that **efficient networks need not be stable and stable networks need not be efficient**. The gap between efficiency and stability is the defining tension of network formation theory, and it has structural parallels throughout game theory (analogous to the gap between Nash equilibrium and social optimum in standard games). This node develops the Jackson–Wolinsky framework, the connections model, pairwise stability, efficiency, and the stability–efficiency gap.

Formal Definition

A **network formation game** is specified by:

- A set of players $N = \{1, \ldots, n\}$.
- A **network** (undirected graph) $g \subseteq \binom{N}{2}$, where $\binom{N}{2}$ is the set of all unordered pairs. Write $ij \in g$ if the link between i and j is present.
- A **value function** $v : \mathcal{G} \to \mathbb{R}$, where $\mathcal{G} = 2^{\binom{N}{2}}$ is the set of all networks on N. $v(g)$ is the total surplus generated by network g.
- An **allocation rule** $Y : \mathcal{G} \times \mathcal{V} \to \mathbb{R}^n$ specifying how the surplus is distributed: $Y_i(g, v)$ is player i's payoff from network g under value function v, with $\sum_i Y_i(g, v) = v(g)$.

Pairwise stability (Jackson–Wolinsky). A network g is **pairwise stable** if: 1. **No player wants to sever a link:** for every $ij \in g$, $Y_i(g, v) \geq Y_i(g - ij, v)$ and $Y_j(g, v) \geq Y_j(g - ij, v)$. 2. **No pair wants to form a link:** for every $ij \notin g$, if $Y_i(g + ij, v) > Y_i(g, v)$ then $Y_j(g + ij, v) < Y_j(g, v)$.

Link formation requires bilateral consent (both endpoints must benefit or be indifferent). Link deletion is unilateral (either endpoint can sever). This asymmetry — **bilateral formation, unilateral deletion** — is the defining feature of pairwise stability and distinguishes it from Nash equilibrium of a simultaneous link-announcement game.

Efficiency. A network g^* is **strongly efficient** if $v(g^*) \geq v(g)$ for all $g \in \mathcal{G}$. It maximizes total surplus without regard to distribution.

The connections model (Jackson–Wolinsky). Players benefit from both direct and indirect connections, with benefits decaying in distance. For parameters $\delta \in (0, 1)$ (decay) and $c > 0$ (cost per link):

$$Y_i(g) = \sum_{j \neq i} \delta^{d_{ij}(g)} - c \cdot |\{j : ij \in g\}|,$$

where $d_{ij}(g)$ is the shortest-path distance between i and j in g (infinite if disconnected, contributing 0).

Intuition

The move from "game on a fixed graph" to "game that forms the graph" is the game-theoretic analogue of the move from "optimization on a fixed domain" to "optimization over domains." In 8.1–8.4, we asked: given this network, what is the equilibrium? Now we ask: what network will rational agents build?

The Jackson–Wolinsky connections model captures the fundamental tradeoff in network formation: **direct links are costly but indirect links are valuable.** Forming a link costs c to each endpoint, but provides access to all nodes reachable through the network, with decaying benefit. A player at the center of a star pays for many direct links but benefits enormously from being close to everyone. A player at the periphery pays for one link but benefits from the indirect connections it provides.

The stability concept — pairwise stability — is weaker than Nash equilibrium. It only checks one-link deviations (add or delete a single link), not multi-link deviations. This makes it tractable but leaves open the possibility that a pairwise-stable network is vulnerable to coordinated multi-link changes. Various strengthened concepts (pairwise Nash stability, strong stability) exist but are harder to work with.

The central result of Jackson–Wolinsky is the **tension between stability and efficiency**: the efficient network (maximizing total surplus) may not be pairwise stable (individual players want to deviate), and the pairwise-stable network may not be efficient. This tension arises because the value of a link depends on the entire network (through indirect connections), but the cost is borne by the two endpoints. Externalities — the fact that forming a link ij affects the payoffs of third parties through changed path lengths — are the structural source of the gap.

Theorem (Jackson–Wolinsky, 1996). No allocation rule Y that is anonymous and component-balanced simultaneously guarantees that all efficient networks are pairwise stable (for all value functions).

This impossibility result is the network-formation analogue of the first welfare theorem's failure in the presence of externalities. It says: you cannot design a "fair" way to share surplus such that the socially optimal network always emerges from decentralized link-formation decisions.

Structural Role

8.5 endogenizes the graph that was exogenous in 8.1–8.4. It is the mechanism-design layer of Module 8: instead of taking the network as given, we ask what networks arise from strategic interaction and what the welfare consequences are.

- **From 8.1:** the graph-game framework, now with the graph as a strategic variable.
- **From 8.4:** public-goods equilibria depend on the graph; network formation determines which graphs are available.
- **To 8.6:** the stability concept of 8.5 interacts with the complementarity structure of 8.6 — on supermodular games, stable networks tend to be denser.
- **To Module 6:** the stability–efficiency gap parallels the incentive-compatibility constraints in mechanism design.

Mathematical Development

The connections model: efficient and stable networks

Efficient networks. Jackson and Wolinsky characterize efficient networks in the connections model:

- If $c < \delta - \delta^2$: the **complete graph** K_n is efficient. Links are cheap enough that every direct connection is worth forming.
- If $\delta - \delta^2 < c < \delta$: the **star** $K_{1,n-1}$ is efficient. Direct links are moderately costly, so it is efficient to route through a hub.
- If $c > \delta$: the **empty graph** is efficient. Links cost more than the direct benefit.

Pairwise-stable networks. Under the equal-split allocation rule $Y_i(g) = \sum_{j \neq i} \delta^{d_{ij}} - c \cdot d_i$:

- If $c < \delta - \delta^2$: K_n is pairwise stable (efficient and stable coincide).
- If $\delta - \delta^2 < c < \delta$: the star is efficient but **not necessarily pairwise stable**. The center of the star pays $(n-1)c$ in link costs but gets $(n-1)\delta$ in direct benefits. For the center, the star is attractive iff $\delta > c$. But peripherals might want to form links among themselves: peripheral i forming a link to peripheral j gains $\delta - \delta^2$ (replacing an indirect connection of length 2 with a direct one). So if $c < \delta - \delta^2$, peripherals want to add links, and the star is not stable. In the intermediate range $\delta - \delta^2 < c < \delta$, the star is stable.
- If $c > \delta$: the empty graph is pairwise stable.

The gap. The tension appears in the regime $\delta - \delta^2 < c < \delta$ with n large. The star is both efficient and stable. But consider an allocation rule that gives the center a smaller share of the surplus. Then the center may want to sever links, making the star unstable even though it is efficient. Jackson–Wolinsky show that no anonymous, component-balanced allocation rule can guarantee stability of all efficient networks across all value functions.

Pairwise stability vs Nash equilibrium

In the **strategic link-formation game** (Myerson, 1991), each player simultaneously announces a set of links they are willing to form: $L_i \subseteq N \setminus \{i\}$. Link ij forms iff $j \in L_i$ and $i \in L_j$. A Nash equilibrium of this game is a profile $(L_1, \ldots, L_n)$ where no player can improve by changing their announcement.

Every pairwise-stable network can be supported as a Nash equilibrium of the link-formation game (with appropriate strategies). But Nash equilibria can also include networks that are not pairwise stable — for example, the empty graph is always a Nash equilibrium (if no one announces any links, no one can improve by unilaterally announcing, because link formation requires bilateral consent). This multiplicity is why pairwise stability, despite being weaker, is the standard concept.

Network formation with transfers

If players can make side payments when forming links, the efficiency–stability gap can sometimes be closed. A link ij forms with transfer t_{ij} from i to j such that both are at least as well off. With unrestricted transfers and complete information, efficient networks can always be supported as pairwise stable (Bloch and Jackson, 2007). The practical limitation is that transfers require enforceable contracts, which may not exist in informal network settings.

Dynamic network formation

Consider a dynamic model: at each period, a random pair (i, j) is selected and can form or sever a link. The process converges to a stochastically stable network — the network most robust to random perturbations. Jackson and Watts (2002) show that stochastically stable networks are always pairwise stable, but not every pairwise-stable network is stochastically stable. The stochastic-stability refinement selects among pairwise-stable networks.

Worked Example

The connections model on 4 nodes

Four players, $N = \{1, 2, 3, 4\}$. Parameters: $\delta = 0.6, c = 0.2$. Since $c = 0.2 < \delta - \delta^2 = 0.6 - 0.36 = 0.24$, we are in the regime where the complete graph is efficient.

Complete graph K_4. Each player has 3 direct links. Payoff:

$$Y_i = 3(0.6) - 3(0.2) = 1.8 - 0.6 = 1.2.$$

Total: $4 \times 1.2 = 4.8$.

Star with center 1. Center payoff: $Y_1 = 3(0.6) - 3(0.2) = 1.2$. Peripheral payoff: $Y_2 = 0.6 + 2(0.36) - 0.2 = 0.6 + 0.72 - 0.2 = 1.12$. Total: $1.2 + 3(1.12) = 4.56 < 4.8$.

Pairwise stability of K_4. Delete link 12: player 1 goes from payoff 1.2 to $2(0.6)+0.36-2(0.2) = 1.56-0.4 = 1.16 < 1.2$. Player 1 does not want to delete. Similarly for all links. Add a link? K_4 is already complete. So K_4 is pairwise stable.

Now change to $c = 0.3$. We have $c = 0.3 > \delta - \delta^2 = 0.24$ but $c = 0.3 < \delta = 0.6$. The star is efficient.

Star payoff. Center: $Y_1 = 3(0.6) - 3(0.3) = 0.9$. Peripheral: $Y_2 = 0.6 + 2(0.36) - 0.3 = 1.02$. Total: $0.9 + 3(1.02) = 3.96$.

Complete graph payoff. Each: $Y_i = 3(0.6) - 3(0.3) = 0.9$. Total: $3.6 < 3.96$. Complete graph is not efficient.

Is the star pairwise stable? Peripheral i considers adding link to peripheral j: marginal benefit $\delta-\delta^2 = 0.24$, cost $c = 0.3$. Net: $-0.06 < 0$. No peripheral wants to add a link. Center considers severing a link: drops from payoff 0.9 to $2(0.6) + 0.36 - 2(0.3) = 1.56 - 0.6 = 0.96 > 0.9$. Wait — the center *gains* from severing a link. So the star is **not** pairwise stable at this allocation.

Rechecking: center with 3 links gets 0.9. With 2 links: $2(0.6) + 0.36 - 2(0.3) = 1.2 + 0.36 - 0.6 = 0.96$. With 1 link: $0.6 + 0 + 0 - 0.3 = 0.3$. The center prefers 2 links to 3. But the severed peripheral loses access. The pairwise-stable network in this range is a "partial star" or path-like structure, not the full star. **This is the stability–efficiency gap in action**: the efficient star is not pairwise stable because the center bears too much cost under equal-split allocation.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Network g	The current graph of study partnerships, coaching relationships, Discord memberships
Link ij	A bilateral study or information-sharing relationship between players i and j
Link cost c	Time, effort, and potential information leakage of maintaining a study relationship
Decay δ	Degradation of information value as it passes through intermediaries
Pairwise stability	Neither partner wants to end the study relationship, and no new pair wants to form one
Efficient network	The graph that maximizes total poker knowledge and edge across the community

Formal object	Poker instantiation
Stability–efficiency gap	Individually rational study partnerships may not produce the best collective outcome

Concrete situation: study group formation

Five mid-stakes regulars live in the same city. Each could pair up with any other for study sessions. A study partnership costs time (c) but provides direct benefits (shared solver output, hand discussion) plus indirect benefits (access to your partner's other partners' insights, decayed by one hop).

The efficient network might be a star — one central organizer who connects everyone — because it minimizes total links while keeping everyone within 2 hops. But the center bears $4c$ in link costs while each peripheral bears only c. Under equal effort-sharing, the center may refuse to be the hub — the stability–efficiency gap.

In practice, poker study networks tend to be **small dense clusters** rather than stars, because the cost of maintaining a study relationship is shared and the benefit is direct and high. This matches the predictions of the connections model in the low-c regime where the complete graph (within a small group) is both efficient and stable. The groups stay small (5–8 people) because adding more members raises cost (more sessions, more coordination) faster than it raises benefit (diminishing marginal knowledge from additional partners).

Concrete situation: coaching market as network formation

Consider the coaching market. A coach (h) and a client (p) form a link if both benefit: the coach gets paid ($Y_h > 0$), the client gets value ($Y_p > 0$). The network of coaching relationships is pairwise stable if no coach wants to drop a client, no client wants to drop a coach, and no unmatched coach-client pair would both benefit from forming a link.

In practice, top coaches limit their client count (they sever links when the marginal time cost exceeds the marginal revenue). Clients leave coaches when they feel the benefit has plateaued (they sever links when Y_p drops below c). New coach-client links form when a referral creates a match that both parties find profitable. The resulting coaching network is sparse, hierarchical (a few high-degree coaches connected to many low-degree clients), and pairwise stable in the Jackson–Wolinsky sense.

The efficiency question is: **is the coaching market's pairwise-stable network also efficient?** Probably not. The efficient coaching network might involve more peer coaching (dense clusters of similar-skill players coaching each other at zero monetary cost), but peer coaching links are harder to form (bilateral benefit is less clear) and easier to sever (no financial commitment). The stability–efficiency gap here is between the sparse, hierarchical, pairwise-stable coaching market and the dense, egalitarian, efficient peer-learning network.

What this understanding buys you

Your poker network is a strategic choice, not an accident. Every study partnership, coaching relationship, and Discord membership is a link you formed (or maintained) because the benefit exceeded the cost. The network you are in right now is pairwise stable — otherwise you would have already changed it. But pairwise stability does not mean efficiency. You may be in a locally optimal network that is globally suboptimal.

The hub position is costly. If you are the central organizer of a study group, you bear disproportionate link costs (scheduling, moderating, curating). The stability–efficiency gap says that the hub position may not be individually rational even when it is socially efficient. If you find yourself in this position and feeling it is unsustainable, that is the Jackson–Wolinsky gap in action — the efficient network (star) is not stable under the current allocation of costs.

Transfers close the gap. In the theoretical framework, side payments can make efficient networks stable. In the poker context, this means: if the hub is compensated (the group pays for the organizer's time, or the organizer gets first access to shared content), the efficient star can become stable. Many successful study groups have an explicit or implicit compensation structure for the organizer — this is transfers making the efficient network pairwise stable.

Cross-Links

- **8.1 Graph games** — the framework where the graph is fixed; 8.5 endogenizes it.
- **8.4 Public goods on graphs** — efficiency of public-goods provision depends on the graph, which 8.5 determines.
- **6.1 Mechanism design** — the stability–efficiency gap parallels the incentive-compatibility constraints.
- **5.4 Folk theorem** — repeated interaction can sustain cooperative network structures that are not stable in a one-shot formation game.
- **2.1 Nash equilibrium** — pairwise stability is a bilateral refinement of Nash for network games.
- **Matching theory** — two-sided network formation has structural parallels to stable matching.

Structural Compression

Invariant: In a network formation game, pairwise-stable networks need not be efficient and efficient networks need not be pairwise stable; the gap arises from externalities in link formation (forming a link affects third parties through changed path lengths), and no anonymous, component-balanced allocation rule eliminates the gap for all value functions.

Mechanism: Bilateral link formation (consent required) and unilateral link deletion; value function depending on the entire network through path-length effects; externalities from indirect connections creating a wedge between private and social returns to link formation.

Cross-System Echo: The Jackson–Wolinsky stability–efficiency gap is the network analogue of the **first welfare theorem's failure under externalities** in general equilibrium theory. It is also structurally parallel to the **price of anarchy** in algorithmic game theory (the gap between the Nash equilibrium and the social optimum in routing, congestion, and resource-allocation games). In all three settings, decentralized decision-making fails to achieve the social optimum because agents do not internalize the full externalities of their choices — and the structure of the externality (positive/negative, local/global) determines the direction and magnitude of the gap.

Exercises

1. For the connections model with $n = 3$, $\delta = 0.5$, enumerate all 8 possible networks and compute $v(g)$ and $Y_i(g)$ for each. Identify the efficient and pairwise-stable networks as a function of c.
2. Prove that the empty graph is always pairwise stable when $c > \delta$. Show that it is the unique pairwise-stable network in this regime.
3. Show that under the equal-split allocation rule in the connections model, the complete graph is pairwise stable iff $c < \delta - \delta^2$. Interpret the condition: adding a direct link replaces an indirect connection of length 2, and the net benefit must be positive.
4. Construct a 5-node network that is pairwise stable but not efficient (under the connections model with specific δ, c). Compute the efficiency loss.
5. In the strategic link-formation game (simultaneous announcements), show that the empty graph is always a Nash equilibrium. Explain why this does not contradict pairwise stability.
6. A poker study group of 6 players has link cost $c = 0.15$ and decay $\delta = 0.7$. Determine whether the complete graph, the star, and the cycle are pairwise stable. Which is efficient?

7. **Argue, structurally,** why the stability–efficiency gap is an inherent feature of network formation with externalities, not an artifact of a particular model. Connect this to the general principle that decentralized decisions with externalities are generically inefficient.

Game Theory · Module 8 — Network Games · Node 8.5
Concepts: graph-theory, nash-equilibrium, stability
Prerequisites: 8.1, 8.4
Enables: 8.6

hbar.university — own.your.cognition