CRIMENET: The global network of organized crime

Alvaro Franco Martins, PhD

April 28, 2026

Abstract

CRIMENET is an open-source database mapping the alliances and rivalries between criminal organizations worldwide. An LLM pipeline extracted 1,890 organizations and 3,354 relationships (2,293 alliances, 1,061 rivalries) from 771 Wikipedia articles. The alliance and rivalry networks are analyzed separately. The alliance network is larger and more cohesive (largest connected component covers 85% of active organizations); the rivalry network is sparser and more fragmented but locally denser. Type assortativity is higher for rivalries than for alliances ($r = 0.64$ vs 0.49): organizations fight within their own type more often than they cooperate within it. The deepest alliance shells are dominated by mafias, including the major American crime families; the deepest rivalry shells are dominated by gangs. Mafias are overwhelmingly cooperative: 86% of their edges are alliances. Gangs sit at the opposite end: nearly half of their edges (45%) are rivalries, the highest rivalry share of any type. Across the two layers, rivalry-dominant brokers are almost all gangs, whereas alliance-dominant brokers are type-mixed. The Hells Angels are the single most central organization in the networks, topping every centrality ranking in both networks.

Keywords: criminal networks, network science, signed networks, organized crime, open-source intelligence, large language models

1. Introduction

Organized crime has been mostly studied at the level of individual organizations. Relationships between organizations have not been mapped at scale. Here I present CRIMENET, an open-source database of alliances and rivalries between 1,890 criminal organizations worldwide, extracted from 771 Wikipedia articles with an LLM pipeline.

The complete pipeline (source URLs, extraction scripts, cleanup logic, deduplication dictionaries, and analysis code) is open-source on GitHub.[1] An interactive D3.js visualization of the full network, with alliance/rivalry filtering and organization search, is deployed on my website.[2]

This report is structured as follows: Section 2 describes the data. Section 3 analyzes the networks. Section 4 lists current limitations.

2. Data

2.1 Source collection

I manually compiled a list of 771 Wikipedia article URLs about criminal organizations, aiming for broad coverage across cartels, mafias, gangs, triads, motorcycle clubs, militias, and terrorist groups.

[1] `https://github.com/alvarofrancomartins/CRIMENET`
[2] `https://www.alvarofrancomartins.com/crimenet`

Wikipedia articles are edited continuously, and the generic URL always resolves to the latest revision. To make the dataset reproducible, I recorded a permanent revision URL[3] for each article rather than the generic one. The text used for extraction was fetched at this exact revision, so every organization and relationship in the dataset traces back to the precise text it was extracted from, regardless of later edits.

For each article, I made two calls to the MediaWiki Action API: one for the clean plain-text body and one for the rendered HTML. Two calls were needed because MediaWiki's plain-text endpoint produces well-structured texts but strips out infobox tables, while the HTML endpoint preserves the infobox but yields lower-quality body text. The HTML was parsed with BeautifulSoup to extract the infobox table (aliases, allies, rivals, years active, etc.), which was appended to the article text. Each article's content and versioned URL were saved to disk.

2.2 Network extraction via LLM

Each article was sent to the DeepSeek API with a structured prompt that enforced a fixed output schema. For each article, the LLM extracted:

- **Entities (nodes):** standardized name, aliases, type, contextual description, and time period. The type was constrained to one of nine canonical values (cartel, mafia, gang, motorcycle club, clan, triad, militia, faction, terrorist organization) or "other".
- **Relationships (edges):** source organization, target organization, relationship type (alliance, rivalry, or other), optional detail, context, and time period. The prompt explicitly mapped cooperation to alliance and conflict to rivalry, reserving "other" when the relationship is genuinely neither cooperation nor conflict.

Articles longer than 2,500 words were chunked, with the article's opening paragraph passed as context to every chunk to preserve entity resolution across chunks. Extraction ran in parallel across 50 concurrent workers and returned a JSON object of nodes and edges for each article.

2.3 Cleanup pipeline

All JSON files were merged into a single global network. Then, nodes were deduplicated by name (case-insensitive), with aliases, descriptions, and source article references merged across duplicates. Edges were deduplicated by (source, target, relationship type, detail). Every node and edge retains the Wikipedia URL it was extracted from. A curated dictionary maps known variant spellings to their canonical form (e.g., "Medellin Cartel" $\rightarrow$ "Medellín Cartel", "Hells Angels MC" $\rightarrow$ "Hells Angels", "FARC-EP" $\rightarrow$ "FARC").

Because the prompt enforced canonical types and relationship classifications directly, the cleanup pipeline focuses on work the LLM cannot do per-article:

- **Organization type overrides.** Organizations the LLM mistyped had their type forced to the correct value via a hand-curated mapping.
- **Hand-curated exclusions.** Non-criminal entities that occasionally slipped through extrac-

[3] Example: `https://en.wikipedia.org/w/index.php?title=%27Ndrangheta&oldid=1349918875`

tion (governments, NGOs, political parties) were removed via a curated exclusion list. The same list was used to remove defunct organizations: those with documented dissolution dates and no evidence of continued activity.

- **Generic node filtering.** Umbrella terms referring to categories rather than specific organizations (e.g., "Russian organized crime", "Colombian drug cartels") were removed. Specific organizations with similar names (e.g., "Mexican Mafia", "Albanian Mafia") were preserved.
- **Type-name safety net.** A small fallback map catches occasional non-canonical type names that the LLM emits despite the prompt constraints (e.g., `crime_family` → `mafia`).
- **Chronological edge deduplication.** Edges connecting the same organizations were deduplicated. When several articles described the same pair, or when the relationship between two organizations changed over time, multiple edges accumulate. Therefore, I pass it through the LLM: if every edge has a parseable date, keep the one with the latest end date. If some edges have dates and others don't, ask the LLM to pick the most current. If no edges have dates, ask the LLM to pick the most current.

2.4 Dataset overview

The final dataset contains 771 source articles with full Wikipedia URL provenance, from which I extracted:

- **1,890 entities** across nine organization types: gangs (39.7%), clans (13.5%), mafias (13.4%), motorcycle clubs (10.4%), factions (9.4%), cartels (4.3%), militias (4.2%), triads (2.8%), terrorist organizations (2.2%).
- **3,354 relationships:** 2,293 alliances (68%) and 1,061 rivalries (32%). All edges are undirected.

Because each node and edge retains its source article, I can also count how many distinct articles mention each organization. Some organizations are referenced far more often than others across the source articles. Table 1 lists the 10 most cross-referenced.

Table 1. Top 10 organizations by number of source articles in which they are mentioned.

Organization	Type	Articles
Hells Angels	Motorcycle Club	93
Sinaloa Cartel	Cartel	69
'Ndrangheta	Mafia	63
Cosa Nostra	Mafia	56
Gambino crime family	Mafia	51
Bloods	Gang	49
Crips	Gang	47
Mexican Mafia	Mafia	47
Camorra	Mafia	45
Bonanno crime family	Mafia	39

Hells Angels lead by a wide margin (93 articles), followed by Sinaloa Cartel (69) and 'Ndrangheta (63). The list is dominated by mafias and reflects the same organizations that will later show up at the top of the centrality rankings.

3. Network structure

3.1 The full network

Alliance and rivalry are very different types of relationships. While the rest of this analysis treats them as two separate networks, seeing them combined gives a helpful overview of the overall scale of these organizations and their connections. Figure 1 shows the active landscape of global organized crime as of April 27, 2026.

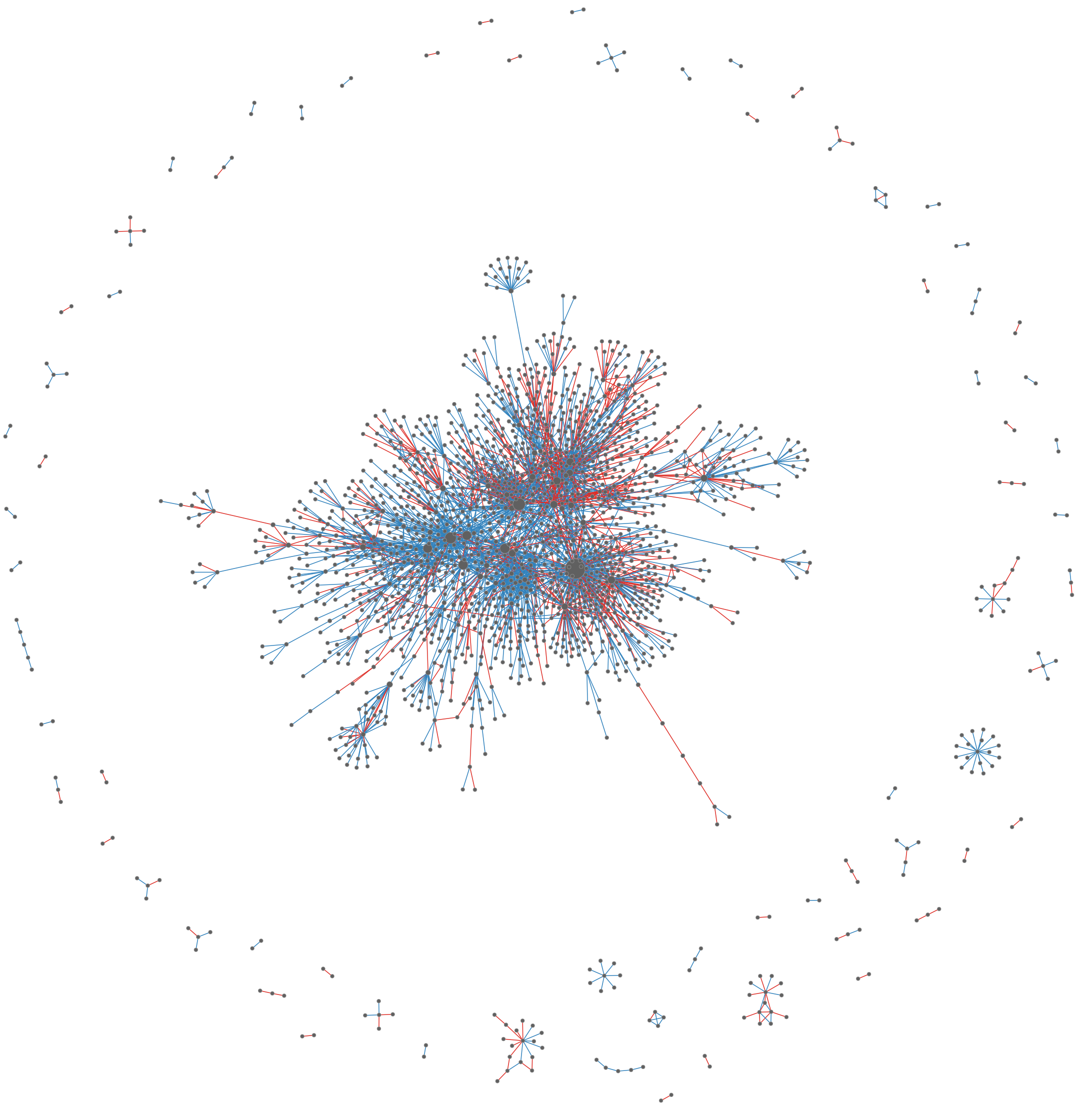

Figure 1. CRIMENET. Nodes are criminal organizations, sized by betweenness centrality. Blue edges are alliances, red edges are rivalries. The entire network contains 1,890 nodes and 3,354 edges. An additional 536 isolated organizations (no extracted connections) are omitted from the visualization (and report).

A clearer picture emerges when each layer is plotted on its own. Figure 2 shows the alliance and rivalry layers side by side.

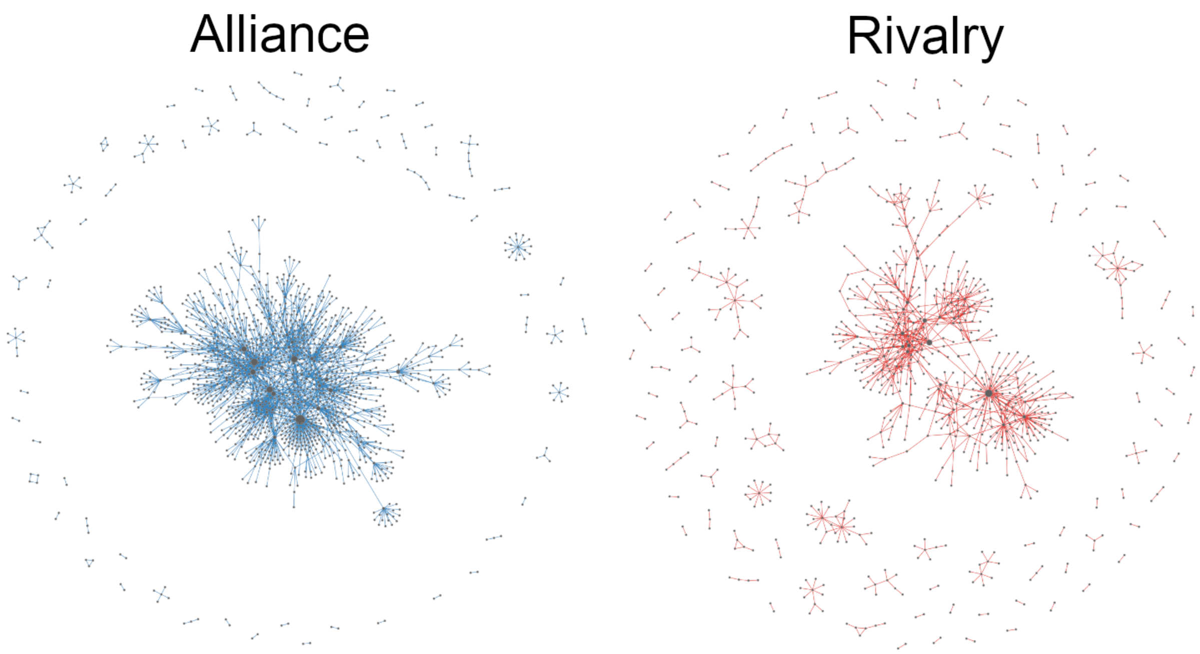

Figure 2. Alliance (left, blue) and rivalry (right, red) networks shown separately.

The first question is a basic one: how do the two networks differ in size, reach, and shape? Table 2 reports the structural properties.

Table 2. Structural properties of the alliance and rivalry networks.

Metric	Alliance	Rivalry
Active nodes (deg > 0)	1,517	876
Edges	2,293	1,061
Density	0.00199	0.00277
Mean degree	3.02	2.42
Median degree	1	1
Max degree	134	53
Components	68	102
LCC nodes	1,292 (85.2%)	478 (54.6%)
LCC edges	2,130 (92.9%)	744 (70.1%)
LCC density	0.00255	0.00653
Pseudo-diameter (LCC)	12	13
Global clustering	0.0906	0.0859
Mean local clustering	0.2957	0.1893
Degree assortativity	-0.1236	-0.0796
Max k-core	8	4
Max betweenness	0.187749	0.134382
Max PageRank	0.027264	0.019312

The alliance network has more active nodes (1,517 vs 876) and edges (2,293 vs 1,061) and its LCC covers 85% of active nodes versus 55%. Cooperation reaches more of the network than conflict does. Global clustering is 0.0906 vs 0.0859, mean local clustering 0.2957 vs 0.1893. Pseudo-diameters are close (12 vs 13). One detail worth noting: the rivalry LCC is denser than the alliance LCC (0.00653 vs 0.00255), even though the rivalry network is sparser overall.

Beyond the overall shape of the networks, it is also useful to measure how often similar groups interact. Type assortativity (Newman's categorical r) is higher for rivalries than alliances ($r = 0.64$ vs 0.48, $p < 0.001$, edge bootstrap). Organizations fight within their type even more often than they ally within it.

3.2 Network centrality

To identify the most prominent organizations in each network, I ranked them using three standard metrics: degree, betweenness, and PageRank. Figure 3 shows the top ten organizations in each network for all three metrics.

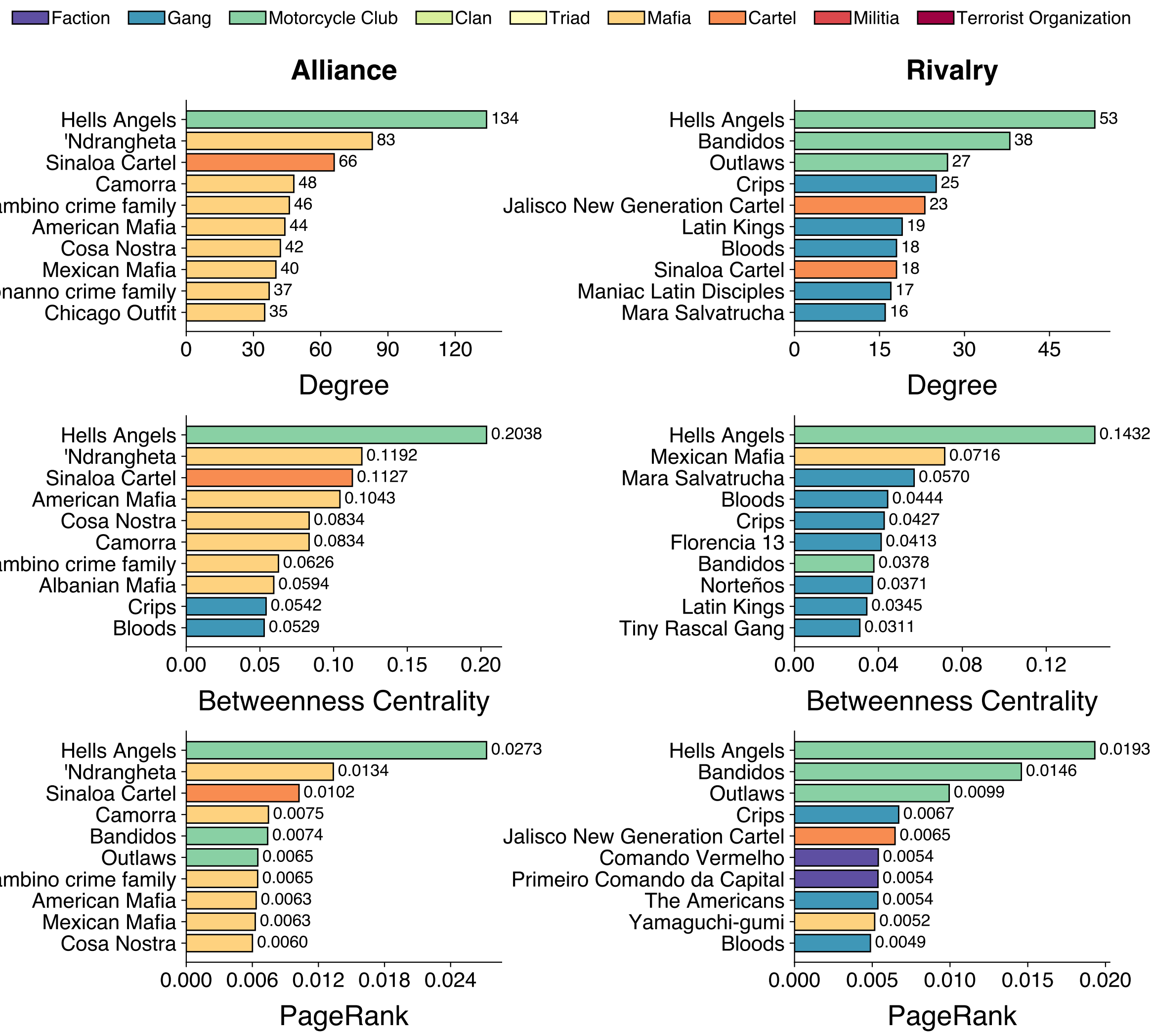

Figure 3. Top 10 organizations by degree (top), betweenness centrality (middle), and PageRank (bottom) for the alliance (left) and rivalry (right) networks. Node colors indicate organizational type.

Hells Angels top every ranking in both networks; no other organization does.

Alliance: Hells Angels, 'Ndrangheta, and Sinaloa Cartel lead, with the rest of the top slots held by mafias (Camorra, Gambino crime family, American Mafia, Cosa Nostra, Mexican Mafia, Bonanno crime family, Chicago Outfit). Interestingly, the only gangs that break into the alliance betweenness top ten are the Crips and Bloods. No other gang reaches the alliance PageRank top ten.

Rivalry: motorcycle clubs move to the front and gangs fill much of the list (Crips, Bloods, Latin Kings, Mara Salvatrucha, Florencia 13, Norteños, Tiny Rascal Gang, Maniac Latin Disciples). Cartels stay prominent (Jalisco New Generation Cartel, Sinaloa Cartel). Notably, the only mafias present are Mexican Mafia and Yamaguchi-gumi.

Alliance and rivalry brokerage are held by mostly different organizations. Since gangs and motorcycle clubs clearly dominate the top of the rivalry network, the next step is to aggregate these metrics to see how each organizational type performs as a whole. Do the patterns hold when averaged across all organizations of a type? Figure 4 shows the mean of each metric by type.

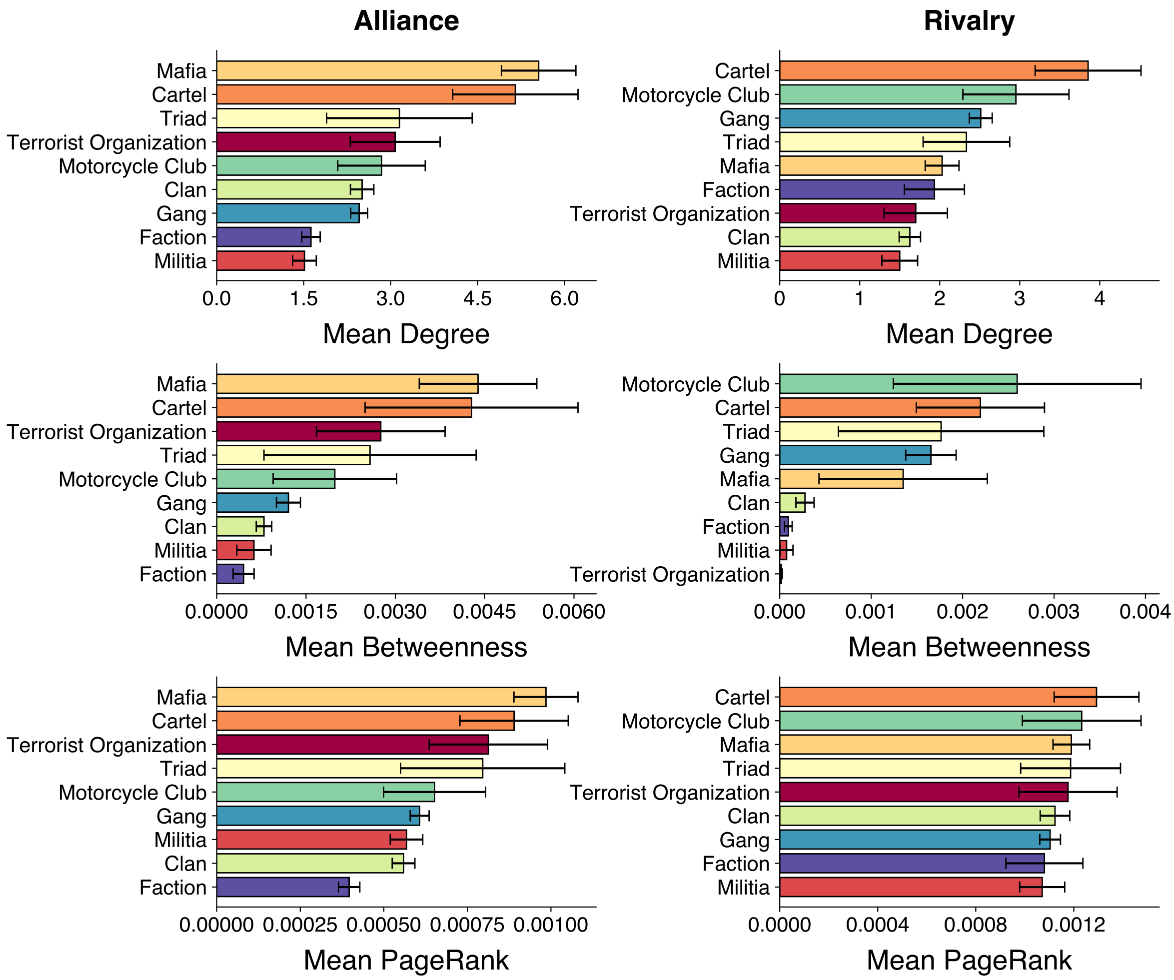

Figure 4. Mean degree (top), betweenness (middle), and PageRank (bottom) by organizational type, with standard error of the mean, for the alliance (left) and rivalry (right) networks.

Mafias and cartels lead the alliance rankings across all three metrics. On the rivalry network, cartels and motorcycle clubs take over and mafias drop. Gangs sit higher in the rivalry ordering than in the alliance one.

3.3 Edge composition by type

Centrality metrics do not reveal the cooperation-conflict mix within each type. For each type, I pool all edges incident on organizations of that type and compute the fraction that are alliances. Figure 5 the result.

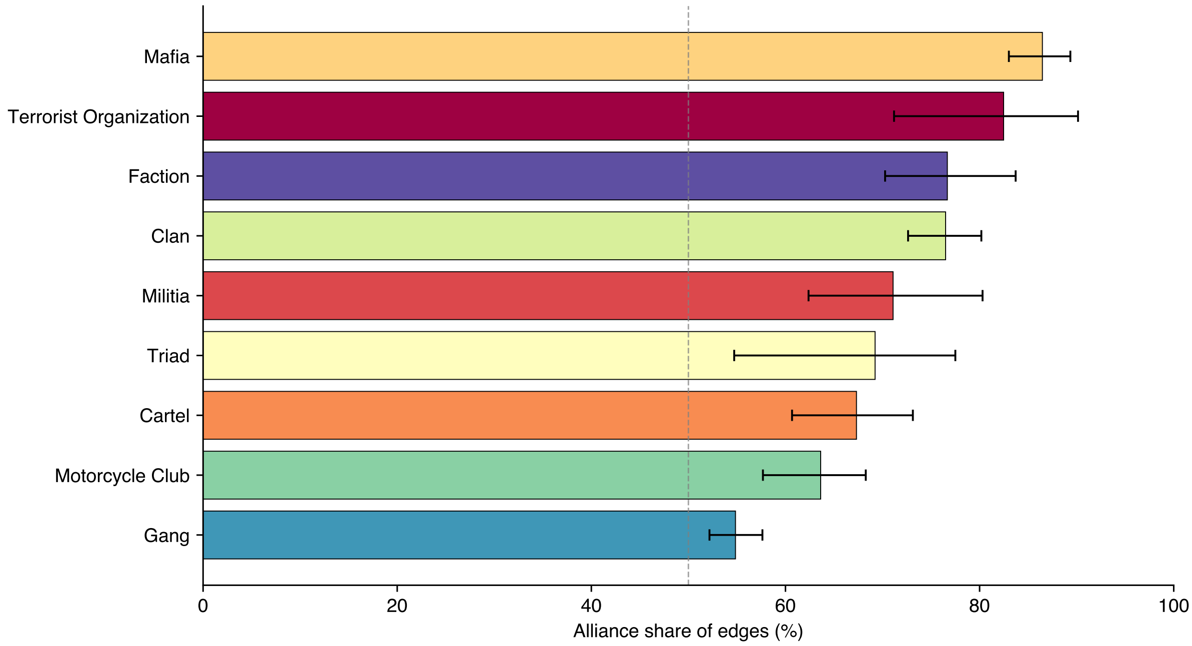

Figure 5. Alliance share of edges by organizational type, edge-weighted (each edge counts equally). Error bars are bootstrap 95% confidence intervals over organizations ($B = 2000$). The dashed line marks 50%.

Mafias spend 86% of their relational footprint on alliances (95% CI [82.5, 89.1]): they are predominantly cooperative. Terrorist organizations, factions, and clans all exceed 75%. Cartels and motorcycle clubs sit in the middle (66% and 64%). Gangs come in last at 55% (95% CI [52.1, 57.5]).

3.4 k-core decomposition

Which types sit at the core of each network? The k-core of a network is the largest subgraph in which every vertex has at least k neighbors. k-core decomposition applies this recursively, peeling off lower-connectivity vertices to reveal a nested shell structure. Figure 6 shows the type composition of each shell.

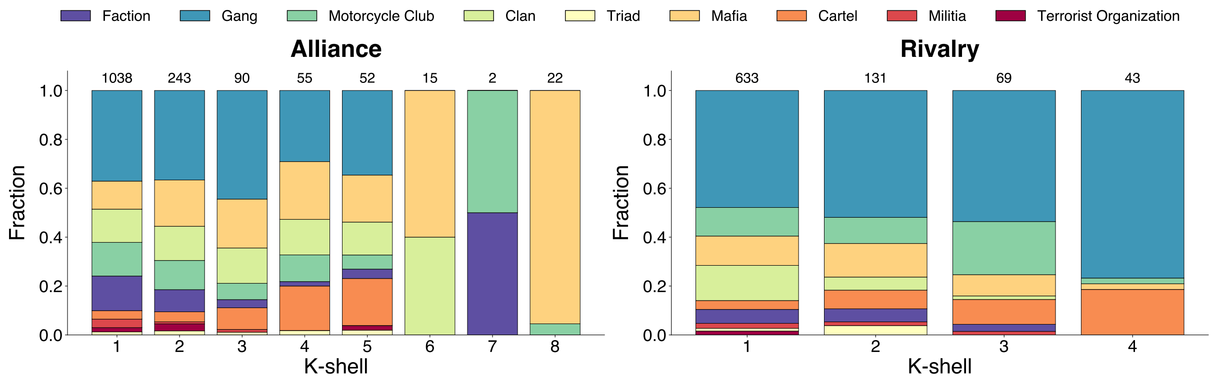

Figure 6. Type composition of k-core shells for the alliance (left) and rivalry (right) networks. Numbers above bars indicate shell size.

The alliance network reaches $k = 8$; the rivalry network only reaches $k = 4$. The alliance core runs deeper. The deepest alliance shell ($k = 8$) contains 22 organizations, almost all mafias—including the major American Mafia crime families (Gambino, Genovese, Bonanno, Lucchese, Colombo), other crime families, and Hells Angels. By contrast, the rivalry $k = 4$ shell contains 43 organizations, predominantly gangs, and gangs dominate every shell of the rivalry network. The rivalry core is shallower and more type-homogeneous shell by shell, consistent with the higher type assortativity reported in Section 3.1.

3.5 Signed pairs

So far I've looked at the two networks separately. Putting them side by side raises a different question: which organizations are pure cooperators, which are pure antagonists, and which keep a mixed book? Table 3 lists the 15 organizations with the highest total signed degree (allies + rivals): the most involved, regardless of direction.

A caveat before reading the table: these counts only reflect relationships documented in the source articles. An organization may have additional allies or rivals that didn't make it into the dataset, either because they weren't mentioned in the relevant Wikipedia articles or because the corresponding article wasn't included in the source list.

Hells Angels lead by a wide margin, with the next ten slots split between cartels (Sinaloa, Jalisco New Generation, Gulf, Los Zetas), motorcycle clubs (Bandidos, Outlaws), and large gangs (Crips, Bloods, Latin Kings, Mara Salvatrucha). Mexican Mafia is the only mafia in the top 15.

Looking at the extremes, the asymmetry between pure cooperators and pure antagonists is striking. The top pure cooperators (only allies, 0 rivals) are almost all mafias: Cosa Nostra (42 allies), Barbaro 'ndrina (22), Detroit Partnership (19), Patriarca crime family (18), Corsican mafia (17), New Orleans crime family (14), Trafficante crime family (14), and the Five Families (13), with two gangs (Jamaican posse, Queen City Kings) as the only non-mafia entries. On the other side, only two organizations qualify as pure antagonists (only rivals, 0 allies): Fresno Bulldogs and Trinitarios, both gangs.

Table 3. Top 15 organizations by total signed degree (allies + rivals).

Name	Type	Allies	Rivals
Hells Angels	motorcycle club	134	53
Sinaloa Cartel	cartel	66	18
Bandidos	motorcycle club	33	38
Outlaws	motorcycle club	33	27
Crips	gang	30	25
Mexican Mafia	mafia	40	14
Bloods	gang	32	18
Jalisco New Generation Cartel	cartel	21	23
Latin Kings	gang	24	19
Gulf Cartel	cartel	30	13
Los Zetas	cartel	28	12
Sureños	gang	29	8
Yamaguchi-gumi	mafia	20	12
Primeiro Comando da Capital	faction	18	12
Mara Salvatrucha	gang	13	16

3.6 Role persistence across networks

Total involvement does not tell us whether the same organizations broker in both layers. If an organization is a broker in the alliance network, is it also a broker in the rivalry network? To find out, I rank each organization active in both layers by betweenness within each network and compare the ranks. Perfect transfer would put points on the diagonal in Figure 7.

Most points fall off the diagonal. To rank persistent brokers, I score each organization by $\max(\text{rank}_a, \text{rank}_r)$, so an organization scores well only if it ranks high in *both* layers, with ties broken by $|\Delta|$. Table 4 lists the top 10.

Table 4. Top 10 most persistent organizations by betweenness centrality (highly ranked in both networks).

Organization	Type	Alliance Rank	Rivalry Rank	$\lvert\Delta\rvert$
Hells Angels	Motorcycle Club	1	1	0
Crips	Gang	8	4	4
Bloods	Gang	9	5	4
Mexican Mafia	Mafia	11	2	9
Bandidos	Motorcycle Club	12	7	5
Outlaws	Motorcycle Club	13	11	2
Sinaloa Cartel	Cartel	3	13	10
Latin Kings	Gang	15	9	6
Los Zetas	Cartel	17	14	3
14K Triad	Triad	14	19	5

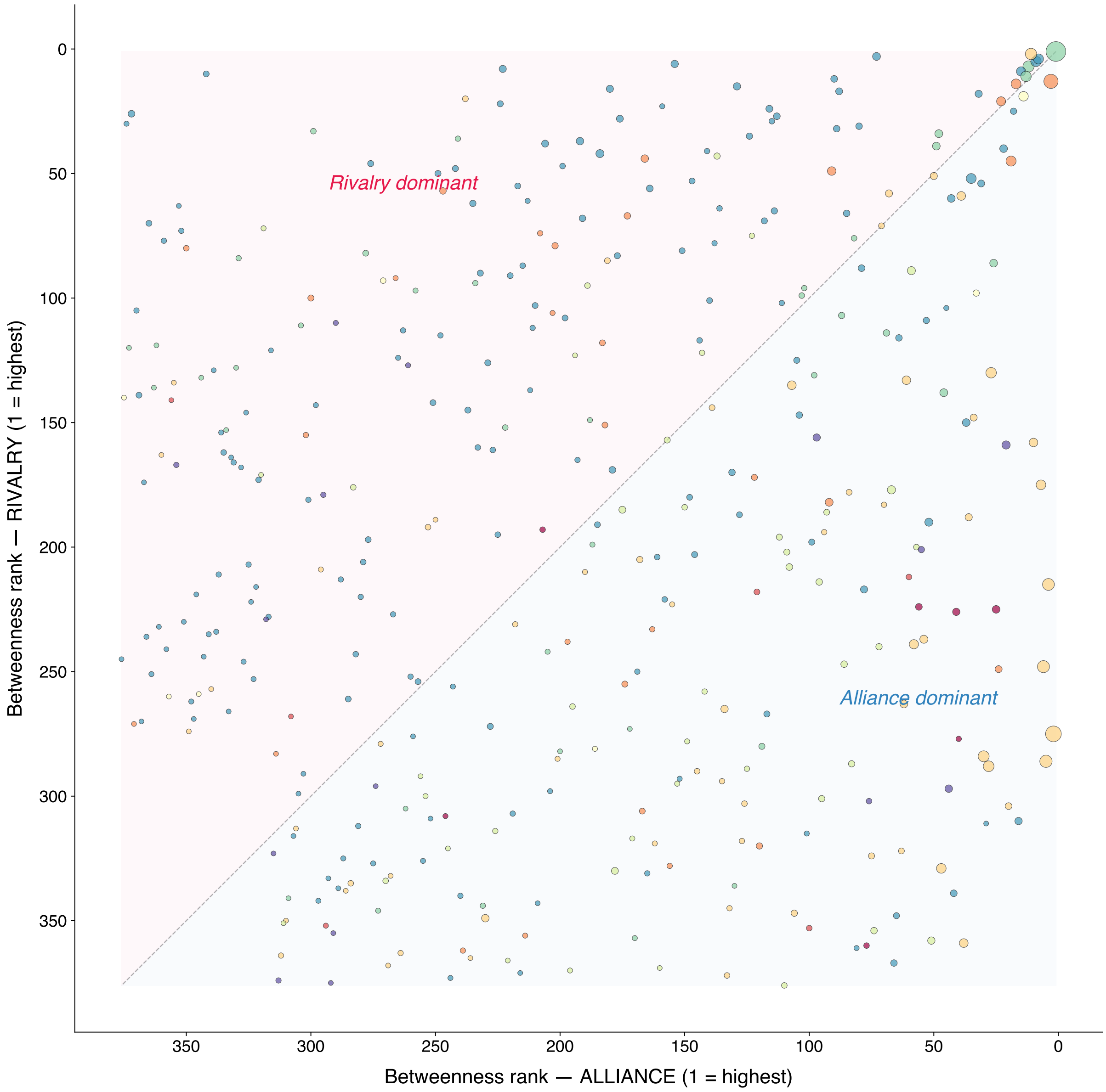

Figure 7. Betweenness rank in the alliance network (x-axis) versus rivalry network (y-axis) for the 376 organizations with nonzero betweenness in at least one layer. Colors indicate organizational type; node size is proportional to degree.

The clearest exception is Hells Angels: rank 1 on both axes, and the system's outlier on every measure I've looked at (top of every centrality ranking in both networks, top of total signed degree).

The points that fall *off* the diagonal split into two groups: alliance-dominant brokers (high alliance rank, low rivalry rank) and rivalry-dominant brokers (the reverse). Tables 5 and 6 list the top 10 in each direction.

Alliance-dominant brokers (Table 5): the list is type-mixed. Three mafias (Serbian mafia, Greek mafia, Lucchese crime family), one clan (Secondigliano Alliance), one terrorist organization (Shining Path), and five gangs appear. What unites them is a relational profile of documented cooperation with little or no recorded conflict.

Table 5. Top 10 alliance-dominant organizations by betweenness rank divergence.

Organization	Type	Alliance Rank	Rivalry Rank	Δ
Serbian mafia	Mafia	38	359	−321
Secondigliano Alliance	Clan	51	358	−307
The British	Gang	66	367	−301
Nazi Lowriders	Gang	42	339	−297
Hard Livings	Gang	16	310	−294
Greek mafia	Mafia	20	304	−284
Piru Street Boys	Gang	65	348	−283
Shining Path	Terrorist Org.	77	360	−283
Hidden Valley Kings	Gang	29	311	−282
Lucchese crime family	Mafia	47	329	−282

Rivalry-dominant brokers (Table 6): the list is overwhelmingly gangs (9 of 10), with one cartel (La Unión Tepito) and one motorcycle club (Original Red Devils MC). Several of the gangs are Los Angeles–area street gangs (Toonerville Rifa 13, West Side Piru, Grape Street Watts Crips, Santa Monica 13, Playboys, Logan Heights).

Table 6. Top 10 rivalry-dominant organizations by betweenness rank divergence.

Organization	Type	Alliance Rank	Rivalry Rank	Δ
Toonerville Rifa 13	Gang	372	26	346
West Side Piru	Gang	374	30	344
Grape Street Watts Crips	Gang	342	10	332
Santa Monica 13	Gang	365	70	295
Los Chone Killers	Gang	353	63	290
Playboys	Gang	359	77	282
Logan Heights Gang	Gang	352	74	278
La Unión Tepito	Cartel	350	80	270
Original Red Devils MC	Motorcycle Club	299	33	266
Tango Blast	Gang	370	105	265

4. Limitations

1. **Wikipedia coverage bias.** Wikipedia skews English-language and Western-centric. Well-documented organizations appear more connected. Coverage gaps exist.

2. **LLM extraction errors.** The sanitization pipeline catches obvious problems, but connections and organization details may still contain errors that survive automated cleanup.

3. **Temporal conflation.** The network aggregates relationships across all time periods. An alliance from the 1970s and one from 2020 have the same weight.

4. **Type classification ambiguity.** The boundary between “cartel” and “militia”, or “gang” and “faction”, is often fuzzy.

5. **Relational oversimplification.** Categorizing complex interactions strictly as alliances or rivalries is an intentional simplification to enable macro-level analysis.