Tie Strength characterization in Dutch Cocaine Trafficking Networks

Keywords:
criminal networks, link prediction, tie strength, network intervention, community structure

Extended Abstract

Criminal networks balance efficiency and security, making the characterization of Granovetter's tie strength essential for understanding their structure and resilience [1]. We study tie strength in Dutch cocaine trafficking networks by analyzing 7,603 edges among 3,341 actors in the giant component of a law-enforcement-derived network. Building on theory and domain expertise, we propose a heuristic tie strength measure that integrates temporal duration, frequency), spatial (proximity), and demographic (age difference, family/social/business ties) features, and we refine it with Random Forest regression [2,3] extensive model selection and validation, duration and spatial proximity emerge as the dominant predictors of tie strength, followed by age difference, while familial and business relationship scores contribute minimally. The regression-based tie strength aligns closely with the heuristic baseline and achieves high explanatory power (adjusted R2 ≈ 0.98). Network intervention experiments show that removing strongest ties first most rapidly reduces clustering, [4] centralities, and giant component size, indicating their role in sustaining local ion; removing weakest ties first increases modularity [5,6], consistent with weak ties bridging communities [1]. We further simulate short-term evolution by adding a +30 day observation to each tie, which increases predicted tie strengths (higher mean, narrower dispersion) and disproportionately elevates initially weak ties. Finally, we find Edge Betweenness Centrality is weakly correlated with estimated tie strength, suggesting it is not a reliable proxy for bridging or functional importance in this clandestine setting [7,8]. Our results provide an empirically grounded, generalizable approach to quantifying tie strength in criminal networks, clarify the complementary roles of weak and strong ties [1], and offer actionable insights for disruption strategies. While this research was performed on real anonymized data, the potential implications for law enforcement mean that the results could inform actual interventions in criminal networks. The researchers are mindful of these ethical considerations and have therefore implemented additional verification and validation measures such as model validation, baseline comparison, network intervention experiments, temporal simulations, and correlation checks with network metrics, to ensure that tie strength predictions are accurate, reliable, and responsibly applied.

References

[1] M. S. Granovetter. The strength of weak ties. American Journal of Sociology, 78(6):1360–1380, 1973.
[2] Leo Breiman. Random forests. Machine Learning, 45(1):5–32, 2001.
[3] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. Bias in random forest variable importance measures: Illustrations, sources, and a solution. BMC Bioinformatics, 8:25, 2007.
[4] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. Nature, 393(6684):440–442, 1998. Accessed: 2024-10-22.

[5] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10):P10008, 2008.
[6] Mark Newman. Networks: An Introduction. Oxford University Press, Oxford, 1st edition, 2010. Online edn, Oxford Academic, 1 Sept. 2010, accessed 14 Aug. 2024.
[7] Ronald S. Burt. Brokerage and closure: An introduction to social capital. Oxford University Press, 2005.
[8] Ronald S. Burt. Structural holes. In Social stratification, pages 659–663. Routledge, 2018.

Table 1. **Feature Dimensions for Tie Strength Calculation.** Overview of the features used to calculate tie strength in Dutch cocaine trafficking networks. Features are grouped into three dimensions: temporal (interaction timing and frequency), spatial (geographic and hierarchical proximity), and demographic (personal and relational attributes).

Feature	Dimension Type
Frequency	Temporal
Duration in Months (DiM)	Temporal
Last Date of Contact (LDoC)	Temporal
Start Date of Contact	Temporal
Hierarchical Distance (HD)	Spatial
Birth Country (PoB)	Spatial
Age Difference in Years (ADiY)	Demographic
Family	Demographic
Markets	Demographic
Gender	Demographic
Business Relationships	Demographic
Family Relationships	Demographic
Combined Roles	Demographic
Family Score	Demographic
Business Score	Demographic
Social Score	Demographic
Active Regions	Demographic

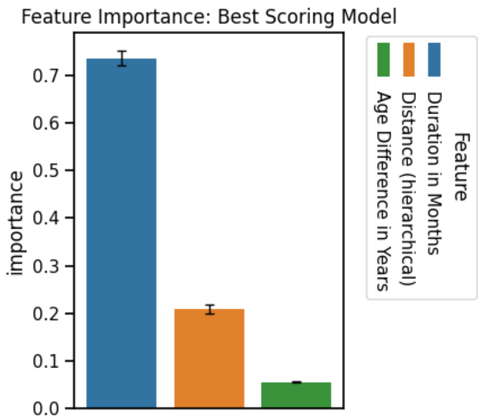

Figure 1. **Best Model Tie Rank Feature Importance's with Error Bars.** Feature importance analysis for the best-performing regression model in predicting tie strength. Mean importance values for each feature, with error bars representing 1.5 times the standard error of the mean (SEM) obtained through bootstrapping (n=100). The numerical values for feature importance and corresponding standard errors are as follows: Duration in Months (0.69 ± 0.0030), Distance (hierarchical) (0.26 ± 0.0025), and Age Difference in Years (0.05 ± 0.0006).