



Faculty of Humanities

Computational Linguistics

Moscow
26 December 2025

Elite Figure Skaters Instagram* Network

Final Project on Social Network Course
Lada Shumakova

* designated as an extremist organization by the Russian government



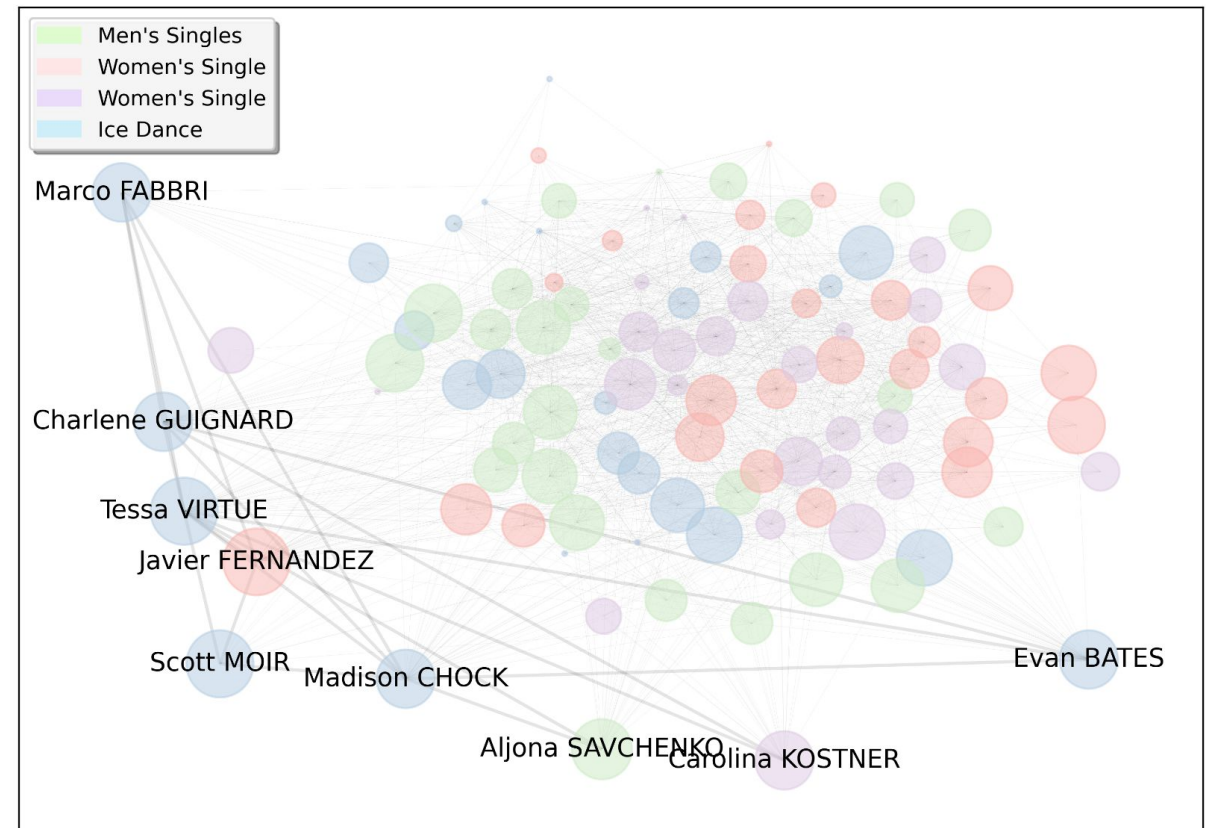
Data Collection & Preprocessing for Figure Skaters' Instagram Network Analysis

- follower relationships among elite figure skaters: `copy_outer_html` to overcome Instagram's dynamic updates
- competition scores scraped from the International Skating Union (ISU) website (HTML Parsing with `BeautifulSoup`)
- sampled 104 skaters equally across all four figure skating disciplines:
 - 26 Women's Singles
 - 26 Men's Singles
 - 26 Pairs (13 men, 13 women)
 - 26 Ice Dance (13 men, 13 women)
- full dataset, preprocessing scripts, and the original notebook are available on https://github.com/vixbi/social_network_project



Network Characteristics

- Structure:
 - 104 nodes representing elite figure skaters' Instagram accounts
 - 2,324 directed edges capturing "following" relationships
- Type properties:
 - directed
 - unweighted
 - homogeneous graph (single professional community)
- Node Attributes:
 - discipline: Women's/Men's Singles, Pairs, Ice Dance
 - Rating Score
 - Full Name: Athlete's official identity.





Summary of Comparison with ER/BA/WS models

Model	Property to be extracted	p-value of KS-test	Interpretation
Erdős–Rényi model	randomness of the network	≈ 0	degree distribution significantly different from random
Barabási–Albert model	scale-free feature presence	≈ 0	the resulting network is not scale-free
Watts-Strogatz	small-world presence	≈ 0	the resulting network may have small world property



Scale-Free Property Analysis

Tested if degree distribution follows a power law using Kolmogorov-Smirnov (KS) goodness-of-fit test → analyzed total degrees (in-degree + out-degree) for directed network.

Empirical Network:

- Optimal lower bound: 49.0
- Power law exponent (α): 4.710
- KS statistic (D): 0.115

Barabási-Albert Model:

- Generated random scale-free network with comparable size/density
- Random network KS statistic: 0.074 (satisfies scale-free criteria)

p-value ≈ 0.000 for two-sample KS-test → **the network is not scale-free**; possible reasons – relatively small size and potential sample imbalance by country/coaching school; missing demographic data limits community analysis depth.



Network Connectivity & Small-World Analysis

Original directed graph not strongly connected → analyzed largest strongly connected component (98 nodes, 2249 edges); performed comparison with Watts-Strogatz Random Model

Metric	Empirical Network	Watts-Strogatz Random Model
Radius	3	2
Diameter	4	3
Avg. Path Length	1.871	1.684
Clustering Coefficient	0.36	0.582

KS test p-value <0.05, average path lengths remain remarkably similar → network exhibits small-world characteristics despite different degree structure; lower clustering → professional networking patterns rather than social friendships



Centrality Analysis: Correlation with Rating Scores

Calculated correlation between ISU rating scores and four network centrality measures:

- Degree Centrality
- Closeness Centrality
- Betweenness Centrality
- Eigenvector Centrality

All Pearson correlations: -0.16 to -0.02 (weak negative) + all p-values > 0.05 → no significant correlation found between network position and athlete rating score.



Interpreting Centrality Patterns

- **Active era factor:** top-central athletes: competed during 2015–2025; least-central athletes: mostly retired
- **Age factor:** younger athletes are more active Instagram users
- **Achievement vs. Connectivity Paradox:**
 - top-central skaters are World/Olympic medalists
 - legendary champions (Hanyu, Kim) have low centrality → elite athletes use Instagram for advertising/branding, not peer networking
- **Career longevity creates hubs:** Elizaveta Tuktamysheva case (top according to all centralities except out-degree centrality)
 - longest active career (2012–2023), competed with both older and younger skaters
 - not the most decorated athlete

Hypothesis: network position reflects active career engagement more than competitive achievement alone



Community Detection Analysis

- **Louvain Algorithm:** efficient for dense social graphs, optimizes modularity to detect overlapping social circles
- **Girvan-Newman Algorithm:** hierarchical edge-removal approach providing alternative structural perspective

Algorithm	Modularity score	Interpretation
Louvain Algorithm	0.185	Weak but detectable community structure
Girvan-Newman Algorithm	0	No meaningful community separation

The Instagram following network exhibits **limited community structure** → professional connections rather than forming tight subgroups; aligns with the small-world characteristics observed previously

