

## 1. Reported Metrics

	ER	BA	Real
N	1000	1000	115
$\langle k \rangle$	4.1420	3.9920	10.6609
$\langle C \rangle$	0.0060	0.0244	0.4032

## 2. Plots

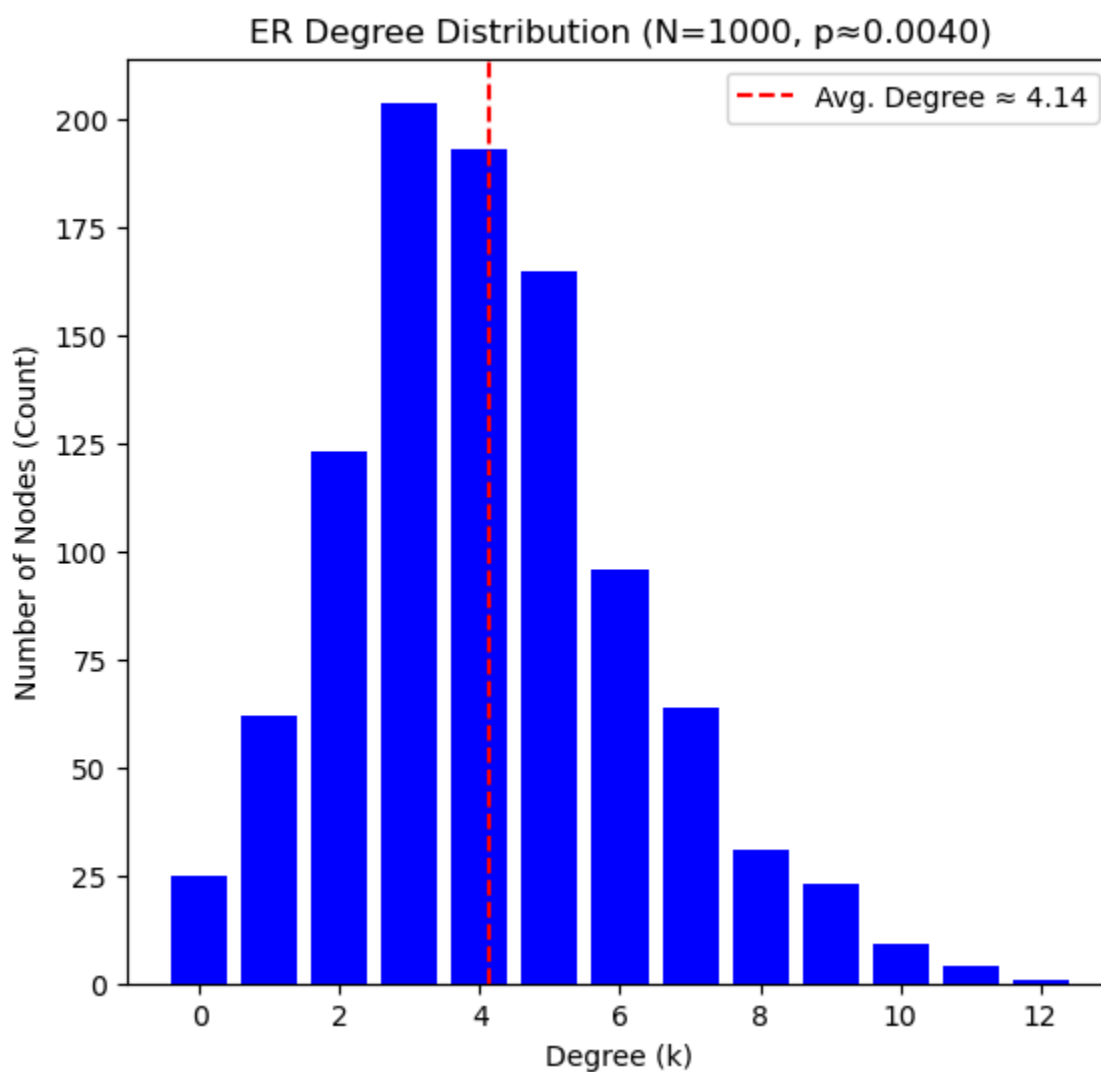


Fig: Bar plot for ER graph

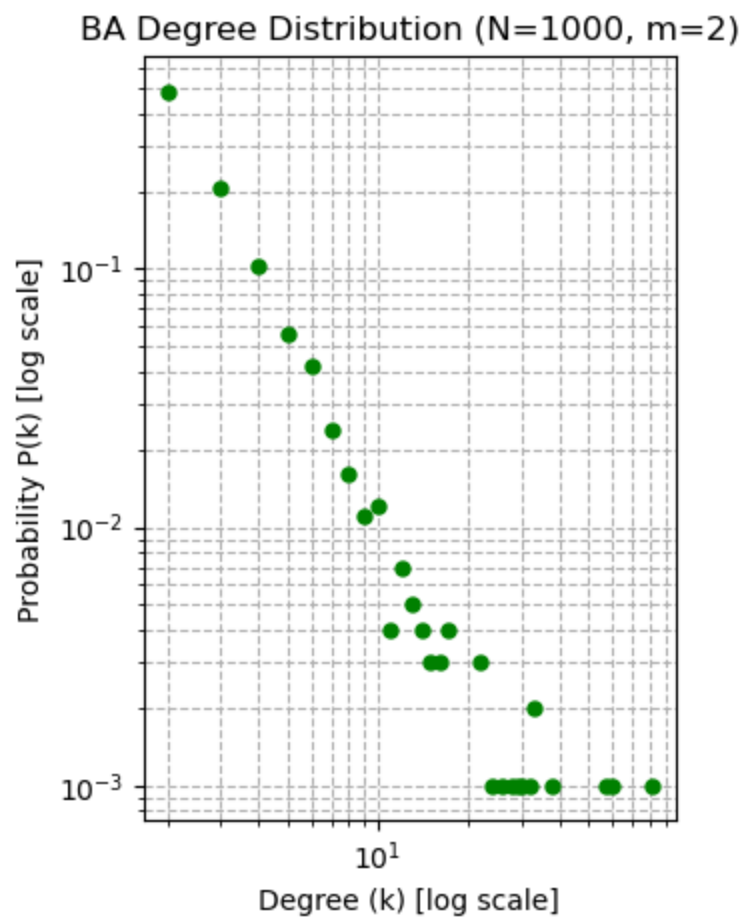


Fig: log-log plot for BA graph

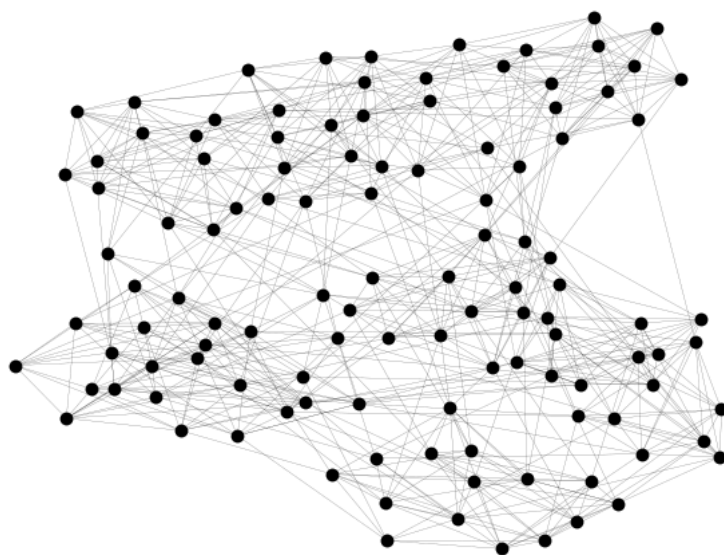


Fig: GML graph for Football Network

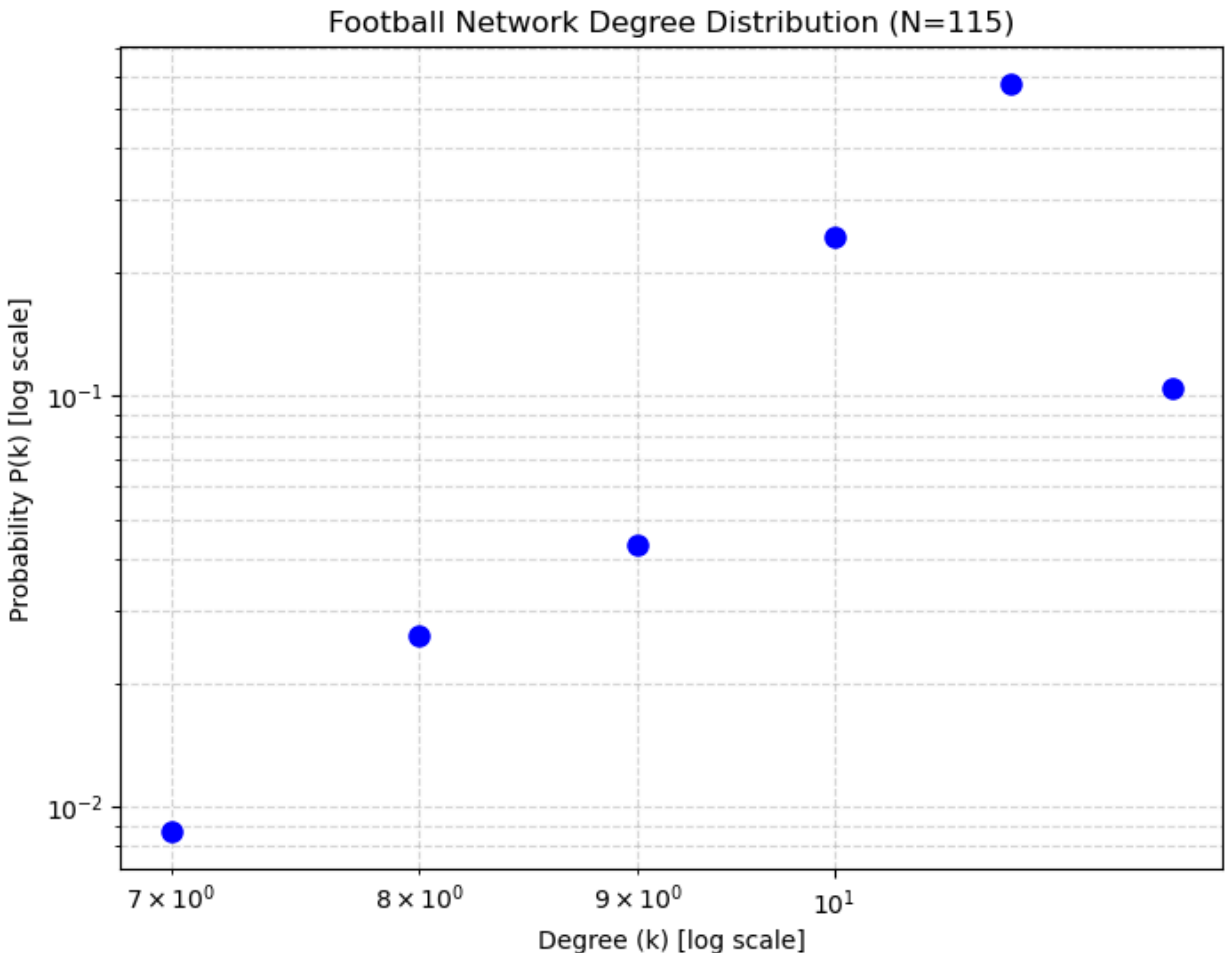


Fig: Degree distribution of the Football Network

### 3. Interpretation

The football network ( $\langle C \rangle \approx 0.403$ ) has the highest clustering by a massive margin, while the Erdos-Renyi (ER) model ( $\langle C \rangle \approx 0.006$ ) has the lowest clustering. The BA model's clustering ( $\langle C \rangle \approx 0.0244$ ) is much higher than the random ER model's, but it's still tiny compared to the real network. This reveals a key weakness of the BA model.

- Why is ER lowest: Connections are purely random. The probability of two of a node's neighbours also being connected (forming a triangle) is minuscule in the BA model.
- Why Football is highest: This network represents a real-world social structure. It's built on communities (the conferences). Teams within the same conference play each other far more often, creating dense, tight-knit clusters. This "friends of my friends are also my friends" effect is called triadic closure, and it's very strong in real networks.
- Why BA is in the middle: The BA model's "preferential attachment" indirectly creates some clustering. When a new node connects to a popular hub, it might also connect to another node that is also connected to that same hub, forming a triangle. However, the BA model has no explicit mechanism for triadic closure. It only models new nodes connecting in, not existing nodes forming new links (e.g., two teams in different

conferences deciding to play because they have a common rival). This is why its clustering coefficient is much lower than a real-world network.

The Barabási–Albert (BA) model is a much better approximation.

- The Football Network's log-log plot (like the BA plot) shows a "long tail," which is characteristic of a power-law distribution. This means there are many teams with a low number of games (low  $k$ ) and a few "hub" teams that play a very high number of games (high  $k$ ).
- The ER model's Poisson distribution is a terrible fit. It predicts that almost all teams would play the average number of games ( $\approx 11$ ) and that hubs are statistically impossible. This is clearly not true for the real network.

The key mechanism responsible for this power-law shape is Preferential Attachment.

- Growth (the network adding nodes over time) is a necessary component, but it's not the whole story.
- Preferential Attachment is the "rich-get-richer" rule: new nodes (or in this case, new games being scheduled) are more likely to connect to nodes that are already well-connected. A well-known "hub" team (like a traditional powerhouse) is more likely to schedule games with other teams than a small, isolated team. This mechanism allows a few nodes to acquire a huge number of connections, creating the power-law tail. The ER model lacks this mechanism entirely, leading to its homogeneous, hub-less structure.