Appendix - A brief research proposal for *leaky pipeline in academia*

Previous research has explored the leaky pipeline phenomenon in academia, which refers to the unequal opportunities and costs faced by male and female researchers along their career trajectories, ultimately leading to the underrepresentation of women in senior academic positions. However, existing studies either use individual-level data (e.g., regional distribution, publication output) to examine gender differences and underlying mechanisms (Wang et al., 2025), without accounting for the potential impact of structural attributes or positional dynamics within collaboration networks; or rely on surveys to measure scholars' integration into academic networks (Leemann et al., 2010). The latter approach may introduce biases from self-reports and lacks granularity in capturing details such as collaboration frequency or the attributes of collaborators. Therefore, this study aims to: a) measure the leaky pipeline phenomenon within academic collaboration networks; b) investigate the potential mechanisms using network data.

Previous research has shown that the level of integration into academic networks is a significant predictor of researcher attrition (Leemann et al., 2010). A promising metric to assess individual integration and activity is the rate of new link formation in dynamic, longitudinal academic collaboration networks. If the rate of link formation predicts node disappearance (i.e., losing all connections in t+1) or changes in node attributes (e.g., academic promotion), research question a can be translated into comparing whether male and female nodes differ in this regard. Barabasi and Albert's (1999) preferential attachment model suggests that new links tend to form with high-degree nodes, and empirical evidence supports a nonlinear relationship between node degree and the probability of link formation in real-world networks (Kunegis et al., 2013). However, this assumption has been challenged in scientific collaboration networks. For example, Abbasi (2016) argued for using structural position features (e.g., closeness and intermediary roles) as alternative predictors of new link formation. Notably, Abbasi defined new links as those between new and existing researchers, while Barabasi et al. (2002) found that links formed among existing researchers contribute most to the preferential attachment model. Therefore, to address research question b comprehensively, this study could evaluate these competing hypotheses to identify which better explains the unequal opportunities women face in forming new links. Lastly, due to the lack of early-stage network data (from the very beginning) and the incomplete nature of snapshots captured at specific time points (with not all researchers included), parameter estimates may be biased (Barabasi et al., 2002). This study proposes using Monte Carlo simulations to compare with empirical results.

The dataset for this study could be SciSciNet (Lin et al., 2023), which contains interrelated data on papers and authors. A supplementary table in this dataset provides gender information based on name classification algorithms, covering 85% of the samples. Additional validation or supplementation could be conducted using tools such as Genderize.io.