Quantifying the impact of weak, strong, and super ties in scientific careers

Alexander M. Petersen
Laboratory for the Analysis of Complex Economic Systems, IMT Lucca Institute for Advanced Studies, Lucca 55100, Italy; Author contact: petersen.xander@gmail.com

A scientist will encounter many potential collaborators throughout the career. As such, the choice to start or terminate a collaboration can be an important strategic consideration with long-term implications. While previous studies have focused primarily on aggregate cross-sectional collaboration patterns, here we analyze the collaboration network from a researcher's local perspective along his/her career. Our longitudinal approach reveals that scientific collaboration is characterized by a high turnover rate juxtaposed with surprisingly frequent 'life partners'. We show that these extremely strong collaborations have a significant positive impact on productivity and citations – the apostle effect – representing the advantage of 'super' social ties characterized by trust, conviction, and commitment. For the long published version, complete with full analysis, references, methods, and data summary, see: AM Petersen *Proc. Nat. Acad. Sci. USA* 112, E4671–E4680 (2015) DOI:10.1073/pnas.1501444112

Here we focus on collaborative endeavors that result in scientific publication, a process which draws on various aspects of social ties, e.g. colocation, disciplinary identity, competition, mentoring, and knowledge flow. The dichotomy between strong and weak ties is a longstanding point of research. However, in 'science of science' research, most studies have analyzed macroscopic collaboration networks aggregated across time, discipline, and individuals. Hence, despite these significant efforts, we know little about how properties of the local social network affect scientists' strategic career decisions.

Against this background, we develop a quantitative approach for improving our understanding of the role of weak and strong ties, meanwhile uncovering a third classification – the 'super tie' – which we find to occur rather frequently. We analyzed longitudinal career data for researchers from cell biology and physics, together comprising a set of 473 researcher profiles spanning more than 15,000 career years, 94,000 publications, and 166,000 collaborators. In order to account for prestige effects, we define 2 groups within each discipline set, facilitating a comparison of top-cited scientists with scientists that are more representative of the entire researcher population (henceforth referred to as "other"). From the N_i publication records spanning the first T_i career years of each central scientists i, we constructed longitudinal representations of each scientist's coauthorship history.

We adopt an ego-centric perspective in order to track research careers from their inception along their longitudinal growth trajectory. By using a local perspective we control for the heterogeneity in collaboration patterns that exists both between and within disciplines. We also control for other career-specific collaboration and productivity differences that would otherwise be averaged out by aggregate cross-sectional methods. Thus, by simultaneously leveraging multiple features of the data – resolved over the dimensions of time, individuals, productivity, and citation impact – our analysis contributes to the literature on science careers as well as team activities characterized by dynamic entry and exit of human, social, and creative capital. Given that collaborations in business, industry, and academia are increasingly operationalized via team

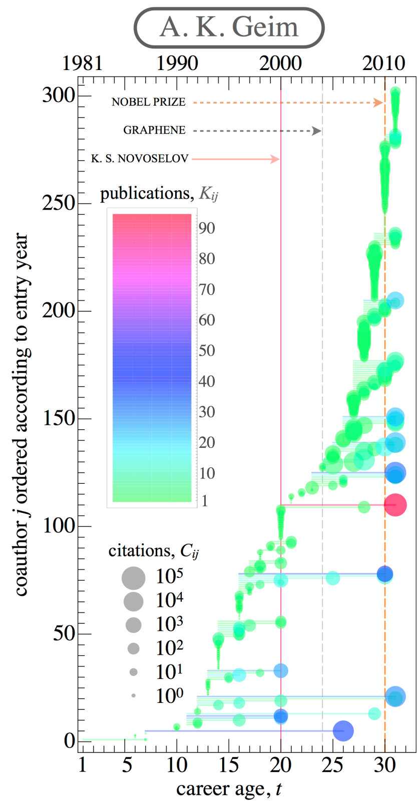

FIG. 1: **Visualizing the embedding of academic careers in dynamic social networks.** A career schematic showing A. Geim's collaborations, ordered by entry year. Each line represents the collaboration between i (Geim) and j lasting L_{ij} years; the circle size represents the cumulative citations C_{ij} from their K_{ij} copublicatons. Notable career events include the first publication in 2000 with K. S. Novoselov (co-winner of the 2010 Nobel Prize in Physics) and their first graphene publication in 2004. An interesting network reorganization accompanies Geim's institutional move from Radboud University Nijmegen (NL) to U. Manchester (UK) in 2001. Moreover, the rapid accumulation of coauthors following the 2004 graphene discovery signals the new opportunities that accompany reputation growth.

structures, our findings provide relevant quantitative insights into the mechanisms of team formation, efficiency, and performance.

In summary, we developed methods to better understand the diversity of collaboration strengths. We focused on the career as the unit of analysis, operationalized by using an 'ego' perspective so that collaborations, publications, and impact scores fit together into a temporal framework ideal for cross-sectional and longitudinal modeling. Analyzing more than 166,000 collaborations, we found that a remarkable 60%-80% of the collaborations last only $L_{ij} = 1$ year. Within the subset of repeat collaborations ($L_{ij} \geq 2$ years), we find that roughly 2/3 of these collaborations last less than a scientist's average duration $\langle L_i \rangle \approx 5$ years, yet 1% last more than $4\langle L_i \rangle \approx 20$ years. This wide range in duration and the disparate frequencies of long and short L_{ij}, together point to the dichotomy of burstiness and persistence in scientific collaboration. Closer inspection of individual career paths signals how idiosyncratic events, such as changing institutions or publishing a seminal study or book, can have significant downstream impact on the arrival rate of new collaboration opportunities and tie formation, e.g. see Figs. 1. We also find evidence indicating that 'career partners' occur rather frequently in science.

In the first part of the study we provide descriptive insights into basic questions such as how long are typical collaborations, how often does a scientist pair up with his/her main collaborator, and what is the characteristic half-life of a collaboration. We also found that as the career progresses, researchers become attractors rather than pursuers of new collaborations. This attractive potential can contribute to cumulative advantage, as it provides select researchers access to a large source of collaborators, which can boost productivity and increase the potential for a big discovery.

We operationalized tie strength using an ego-centric perspective of the collaboration network. We found that the number of publications K_{ij} between the central scientist i and a given coauthor j to be exponentially distributed. Thus, the mean value $\langle K_i \rangle = S_i^{-1} \sum_{j=1}^{S_i} K_{ij}$, calculated across the S_i distinct collaborators (the collaboration radius of i), represents an intrinsic collaboration scale that grows in proportion to both an author's typical collaboration size as well as his/her publication rate. $\langle K_i \rangle$ is a natural author-specific threshold that distinguishes the strong ($K_{ij} \leq \langle K_i \rangle$) from the weak ties ($K_{ij} < \langle K_i \rangle$). Furthermore, within the subset of strong ties, we identified 'super tie' outliers using an analytic extreme-statistics threshold $K_i^c = (\langle K_i \rangle - 1) \ln S_i$ such that if $K_{ij} > K_i^c$ then j is a super-tie of i. On a per-collaborator basis, the fraction of coauthors within a research profile that are super ties was remarkably common across datasets, indicating that super ties occur at a rate of 1 in 25 collaborators, on average.

There are various candidate explanations for why such extremely strong collaborations exist. Prosocial motivators may play a strong role, i.e. for some researchers doing science in close community may be more rewarding than going alone. Also, the search and formation of a compatible partnership requires time and other social capital investment, i.e. networking. Hence, for two researchers who have found a collaboration that leverages their complementarity, the potential benefits of improving on their match are likely outweighed by the long-term returns associated with their stable partnership. Complementarity, and the greater skill-set the partnership brings, can also provide a competitive advantage by way of research agility, whereby a larger collective resource base can facilitate rapid adjustments to new knowledge fronts, thereby balancing the risks associated with changing research direction. After all, a first-mover advantage can make a significant difference in a winner-takes-all credit & reward system.

Scientists may also strategically pair up in order to share costs, rewards, and risk across the career. In this light, an additional incentive to form super ties may be explained, in part, by the benefits of reward-sharing in the current scientific credit system, wherein publication and citation credit arising from a single publication are multiplied across the a_p coauthors in everyday practice. Considered in this way, the career risk associated with productivity lulls can be reduced if a close partnership is formed. For example, we observed a few 'twin profiles' characterized by a publication overlap fraction $f_{K,i}$ between the researcher and his/her top collaborator that was nearly 100%. Moreover, we found that 9% of the biologists and 20% of the physicists shared 50% or more of their papers with their top collaborator. This highlights a particularly difficult challenge for science, which is to develop a credit system which appropriately divides the net credit, but at the same does not reduce the incentives for scientists to collaborate. Thus, it will be important to consider these relatively high levels of publication and citation overlap in the development of quantitative career evaluation measures, otherwise there is no penalty to discourage coauthor free-riding.

We concluded the analysis by implementing two fixed-effects regression models to determine the sign and strength of the 'apostle effect' according to the fundamental question: is there a measurable advantage associated with heavily investing in a select group of research partners?

In the first model we measured the impact of super ties on a researcher's annual publication rate, controlling for career age, average team size, the prior experience of i with his/her coauthors, and the relative contribution of super ties within year t. We found larger *relative* coauthor contributions by super ties to be associated with above-average productivity, indicating that super ties play a crucial role in sustaining career growth. We also found increased levels of prior experience to be associated with decreased productivity, suggesting that maintaining redundant ties conflicts with the potential benefits from mixing new collaborators into the environment.

In the second regression model we analyzed the impact of super ties on the citation impact of individual publications, using the detrended citation measure $z_{i,p,y}$ that is useful for comparing publications (p) from different years y and i. This citation measure is normalized within publication year cohorts, thus allowing for a comparison of citation counts for research articles published in different years. We found that publications coauthored with super ties, corresponding to 52% of the papers we analyzed, have a significant increase in their long-term citations. For example, we show that that publications with super ties receive 17% more citations.