**Group:** Gregg Maloy, Jacob Silver, Justin Williams

For our dataset, we found a [collection of posts](https://www.kaggle.com/datasets/shreyasajal/linkedin-influencers-data?resource=download) from prominent influencers on the professional social media platform LinkedIn in 2021. The data is downloadable from Kaggle, and contains information on the posting influencer, the text of their posts, and additional metadata like the user’s self-reported location and number of connections. The data does not come in a network-readable format and will require pre-processing.

Because LinkedIn thrives on connections, prolific users of the platform often tag others in their posts; LinkedIn themselves [encourage this behavior](https://www.linkedin.com/help/linkedin/answer/a525082/mention-people-in-your-posts?lang=en#:~:text=Mentioning%20a%20connection%20or%20other,about%20the%20post%20or%20comment.), noting that it is likely to increase engagement with a user’s posts, particularly by the tagged individuals. It stands to reason that influencers on the platform would want to engage with other prominent users, thereby aiding visibility and facilitating potential future opportunities.

We can use the author and post text data to create a list of network edges, where the source is a post’s author and the target is any influencer mentioned in the text. While the data is by no means comprehensive of all LinkedIn influencers (only 69 were included in the data collection), any mention-based engagement would still be valuable in understanding how big names on the platform interact.

Further, centrality metrics may tell us something about which influencers are important in particular ways. Degree centrality would indicate who mentions or is mentioned by other influencers prolifically. Betweenness centrality tells us which influencers may serve as “bridges” to different audiences (say, across industries). And we can use eigenvector centrality to see which users are highly influential, even if they themselves are not involved in the most tagging.

Once we calculate these metrics, we can compare their values across different categories. For example, do users posting from a particular city (like New York or London) tend to have more importance in the influencer network? And if so, would it benefit an aspiring influencer to focus his or her efforts on that locality?

It is worth noting that this edge data would be directional, in that tagging someone on LinkedIn is not the same as being tagged. We would need to decide whether to account for this directionality in our analysis, or whether standard centrality metrics could be appropriate.