Taking a long-view of tweeting – an example looking at @HelenBevan's account

Introduction: A few days ago I contacted Helen Bevan to share a social network map of her tweets over a period of almost 2 years. Helen has been a great source of support over the past 5+ years after we met in social media discussions about quality improvement and then in person with the Q Community.

I had run this most recent map of Helen's tweets as an experiment to look at long range tweeting. Usually it is only possible to look at a few days of tweets. However Twitter allows you to extract tweets over a longer period if looking at a single account, providing access to up to 3,200 tweets and retweets. Helen is well known as one of the UK's top healthcare tweeters. She is also very supportive of colleagues from across the world, reading and commenting on others' tweets and blogs and is quick to share useful content with her 86,000+ followers. Mapping her social network connections would help her understand her audience and the content that had most impact. Helen tweeted my map. The map intrigued and confused Helen's followers, so I have posted a blog on this analysis.

Methods: I used NodeXL to extract a "user network" using Helen's Twitter handle – @helenbevan. This downloaded up to 3,200 tweets and retweets made by Helen, and plotted the connections in her original tweets (e.g. Twitter accounts she had mentioned) and the accounts that had been retweeted. The resulting map looks impressive, but when you look at the connections you see that they all lead straight back to Helen rather than connecting with other accounts. In order to understand the wider connections you need to transform the data. I extracted the unique identifier in Helen's tweets and retweets and imported these directly. As long as you have the unique identifier Twitter allows you to download information about that tweet, right the way back to the start of Twitter. It's an

unexpected but useful workaround. When NodeXL imports this information for one of Helen's retweets it downloads data about the original tweet – something that was missing in the original "user network".

NodeXL user map for @HelenBevan
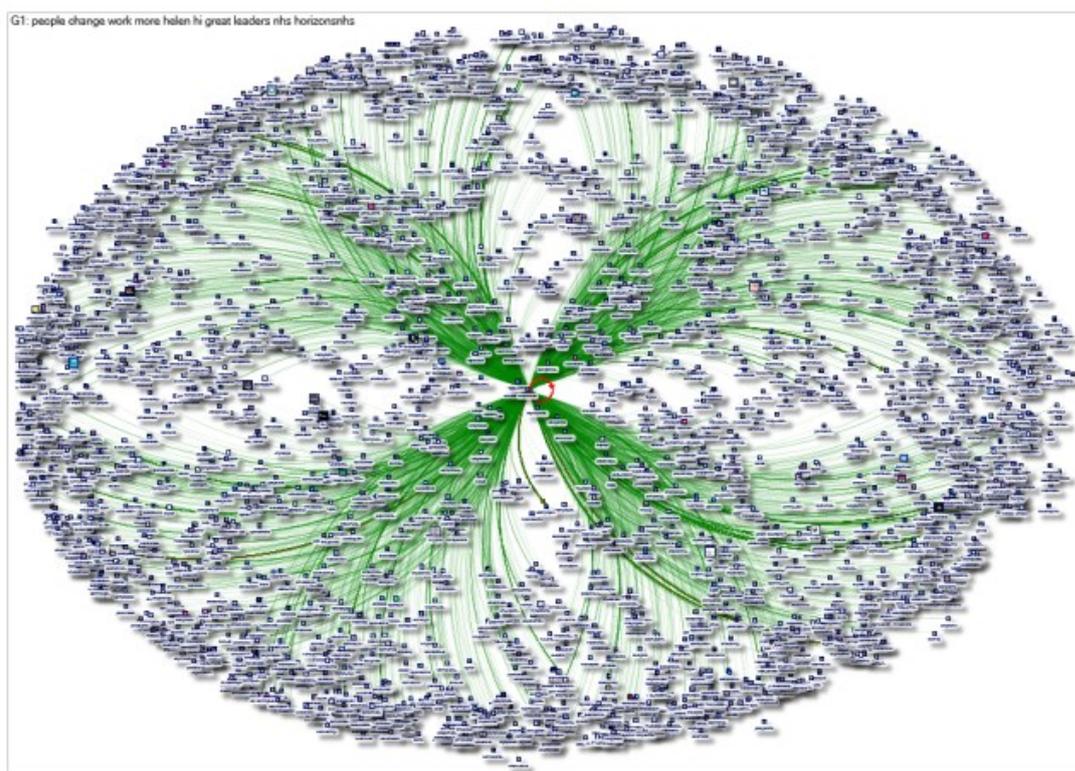
With the full content of the tweet it is able to identify Twitter acccounts mentioned in the same tweet, and maps these connections. This "tweet ID" map is therefore much more useful than the "user network map".

Reconstructed NodeXL map using tweet and retweet IDs
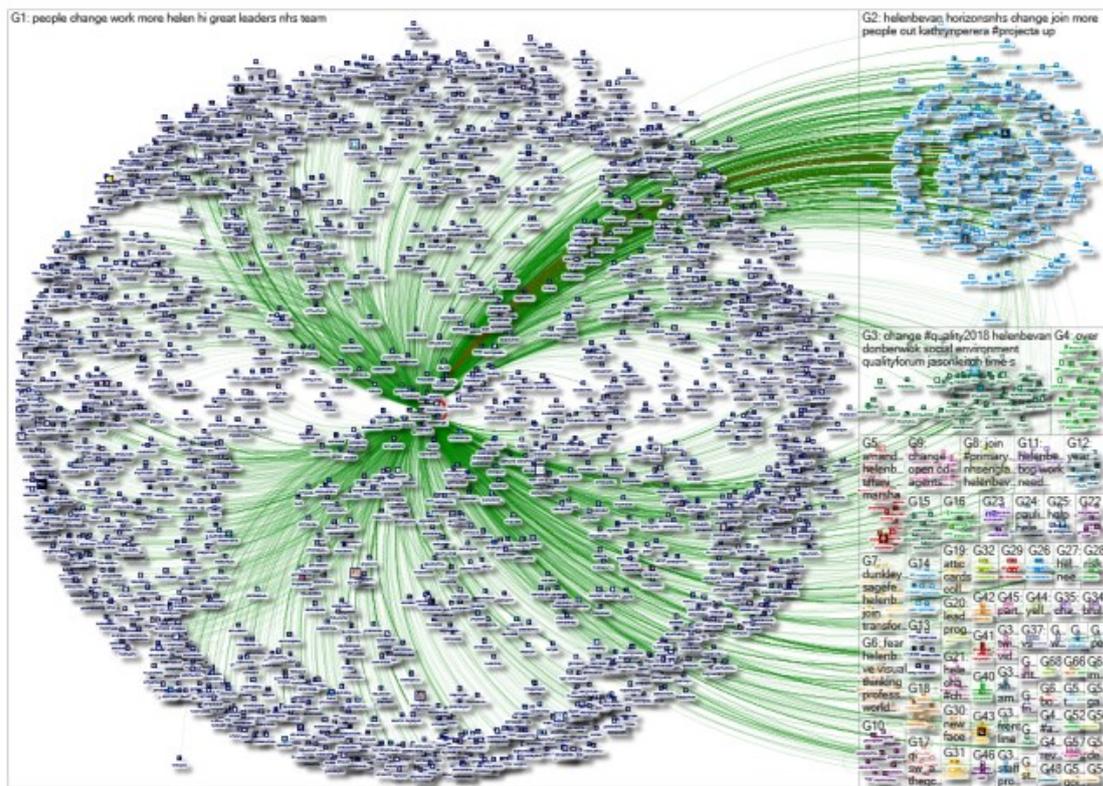
Results: The maps capture data on 3,192 tweets, 86% posted by Helen, 14% retweeted by Helen. The tweets were posted between Wednesday, 22 August 2018 at 18:20 and Saturday, 20 June 2020 at 07:21 UTC. The original tweets were posted by Helen and 280 other tweeters, and in total received 232,531 retweets by the time of the extract on 20 June 2020. If you would like to see the detail then you can see the original in the accompanying Wakelet summary. The green lines in the reconstructed network map show connections and communities. Many of the connections are direct to Helen (G1), but you can also see other communities in the map on the right – e.g. NHS Horizons and connections (G2); the Institute for Healthcare Improvement and associated influential tweeters (e.g. Jason Leitch and Don Berwick (G3). Clusters form because of discussion and replies in tweet exchanges or because tweeters are mentioned together in a tweet. Retweeting is also captured in this map, but mentions and replies in tweets are the most important connections.

We can use other information in the tweets to explore these maps further. Looking just at Helen's original tweets (not replies – which are displayed differently on Twitter), Helen's tweets received a median of 43 retweets (interquartile range 14 to 99 retweets), with the most popular post receiving 2,182 retweets. This is an impressive figure – corporate NHS accounts with larger numbers of followers frequently achieve a much lower number of retweets. Helen's engagement and interaction with other tweeters generates trust, builds audience and encourages wider dissemination of the content that she shares (both tweets and retweets). I have extracted data about the most shared tweets – from Helen and other tweeters – in the Wakelet summary (listed chronologically). This record captures 100 tweets with at least 300 retweets (n=76 tweets by Helen, 24 tweets by other users). These tweets reflect Helen's interests on Twitter – quality improvement, leadership, teams, communication, humour, healthcare events, social media and much more, including COVID-19 more recently. The hashtags used in these tweets are captured in word clouds in the Wakelet summary. One of the word clouds is shown below. The more times the hashtag was used in tweets, the bigger the word in the word cloud. Click the image to see the interactive version of this word cloud, where you can see the individual tweets by clicking on the chosen hashtag.

Conclusions: This blog has described how to take a long view of social media – by focusing on a single Twitter account. I selected Helen Bevan's account because she has a particularly engaging approach to tweeting. The blog has described the two stage approach required to fully understand the connections in these tweets. It has also pulled out the top influencers, hashtags and tweets, as detailed in the Wakelet summary. These approaches have wider applications, but the analysis and interpretation takes time. It also relies on access to the data – for example it is not possible to look back beyond August 2018 using this approach for Helen Bevan's tweets. It is possible to go further back manually, but this would involve much more time. For posterity I have downloaded other healthcare social influencers, starting with international healthcare journals, and hope to post more about that work in due course.

Dr Graham Mackenzie, GPST2, Edinburgh, Scotland

25 June 2020.