

Multimedia Appendix 4

A network of connections between lifestyle-related words (lifestyles, in short) was built and visualized according to the statistics described in the section Analyzing Patients' Tweets. The different lifestyles were the nodes and an arc connected two lifestyles if they appeared in the same tweet. The more times they appeared together, the stronger the connection between the lifestyles was. Therefore, the resulted network was undirected and weighted by the number of times the lifestyles co-occurred. The purpose was to identify helpful lifestyles (frequently mentioned in a positive context) and lifestyles that are better to avoid (frequently mentioned in a negative context) and examine whether certain lifestyles tend to be implemented together.

The network was obtained using *Gephi* software for network analysis and visualization. Each node was colored on a scale from green to red based on the mean sentiment of the lifestyle it represented, green being very positive and red being very negative. The sizing of the nodes reflected the number of times the lifestyles were mentioned in the tweets database: the more times they appeared, the larger their nodes were. The thickness of each arc represented the number of times the two lifestyles it connected co-occurred: the thicker the arc, the more times the two lifestyles appeared together. To avoid obtaining an over-dense network, we only considered nodes of lifestyles mentioned at least five times in our database and included arcs between lifestyles that co-occurred at least four times. The process resulted in 144 lifestyles that are presented in the network.

