

Fantastic Four

Spider-Man

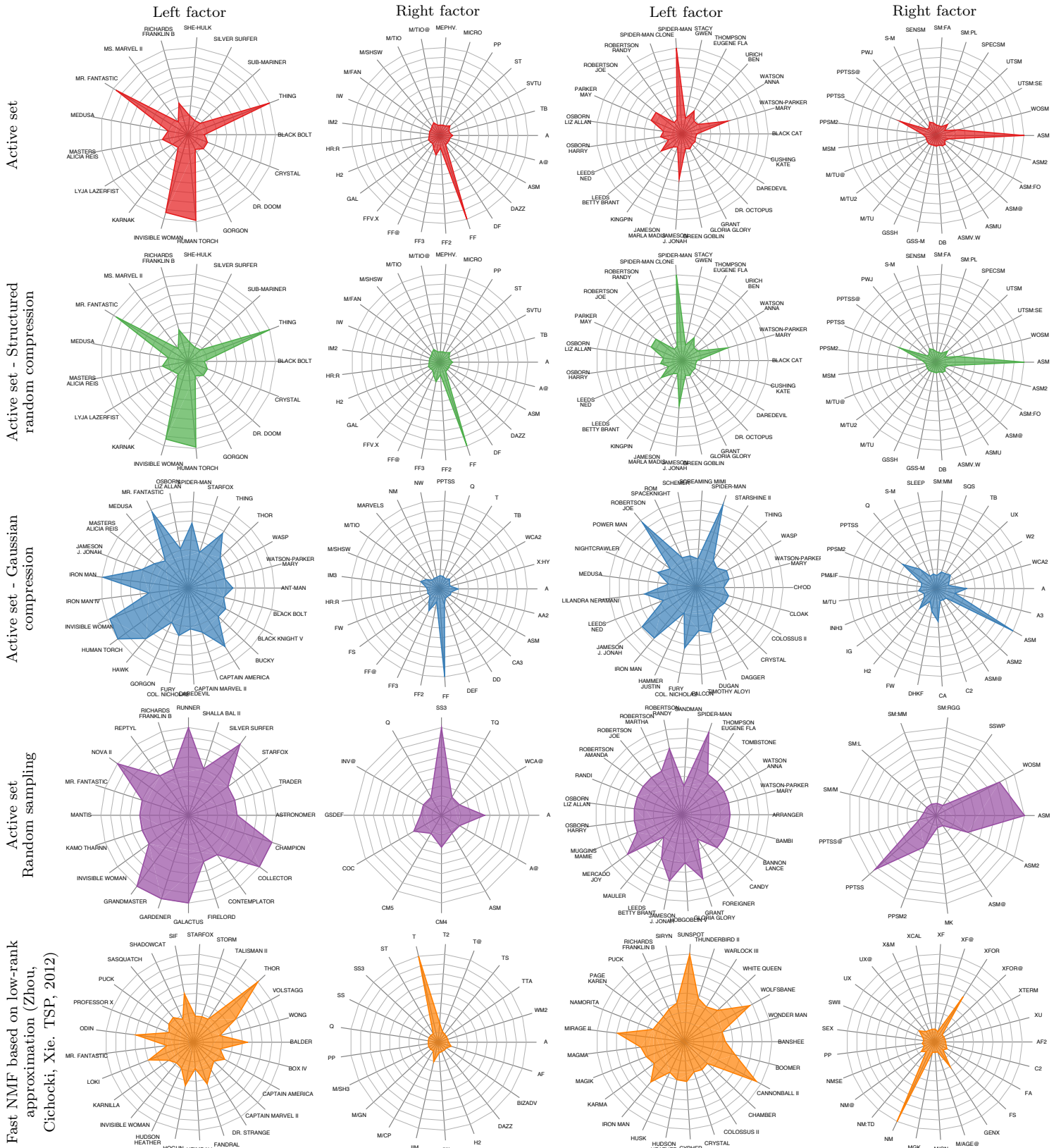


Figure 1. **Biclustering the Marvel Universe collaboration network** (<http://www.chronologyproject.com/>). The network links Marvel characters and the Marvel comic books in which they appear, and exhibits most characteristics of real-life collaboration networks. It can be represented as an  $m \times n$  matrix, where  $m = 6445$  and  $n = 12850$  are the number of characters and comics, respectively. We bicluster this matrix using NMF with  $r = 10$ , aiming at obtaining 10 very representative groups of characters appearing jointly in different comic books. The  $i$ th bicluster ( $i = 1 \dots 10$ ) is formed by the  $i$ th column of  $\mathbf{X}$  and the  $i$ th row of  $\mathbf{Y}$  (small entries were set to zero). The radar plots represent the coefficients of these vectors. We show two biclusters that we identify with characters from the Fantastic Four (first two columns of the figure) and the Spider-Man (last two columns of the figure) comics. Active set NMF correctly identifies that Mr Fantastic, The Thing, the Invisible Woman, and the Human Torch are the four most recurring characters in the “Fantastic Four” (FF) series. Similarly, active set NMF correctly identifies that Spider-Man/Peter Parker, Mary Jane Watson-Parker (Peter Parker’s wife), and Jonah Jameson (Peter Parker’s boss) are the most recurring characters in the “Amazing Spider-Man” (ASM) and “Peter Parker, The Spectacular Spider-Man” (PPTSS) series. It is clear that the biclusters recovered using structured random compression (SC) are very close to the biclusters found with no compression; contrarily, the other methods (GC and RS respectively stand for Gaussian compression and random sampling) significantly affect the biclustering result. All 10 biclusters can be found at <http://www.marianotepper.com.ar/research/cnmf>.