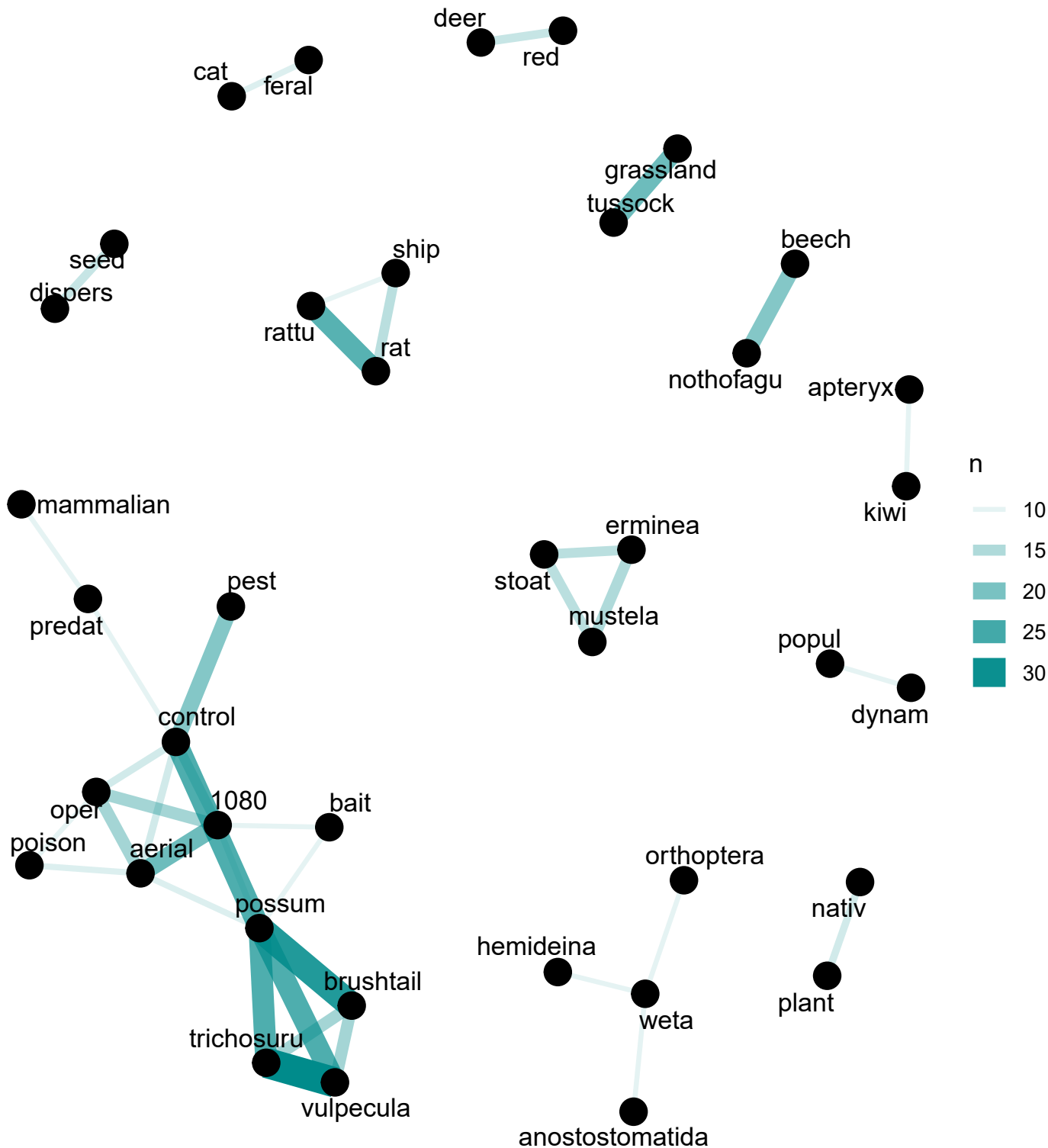
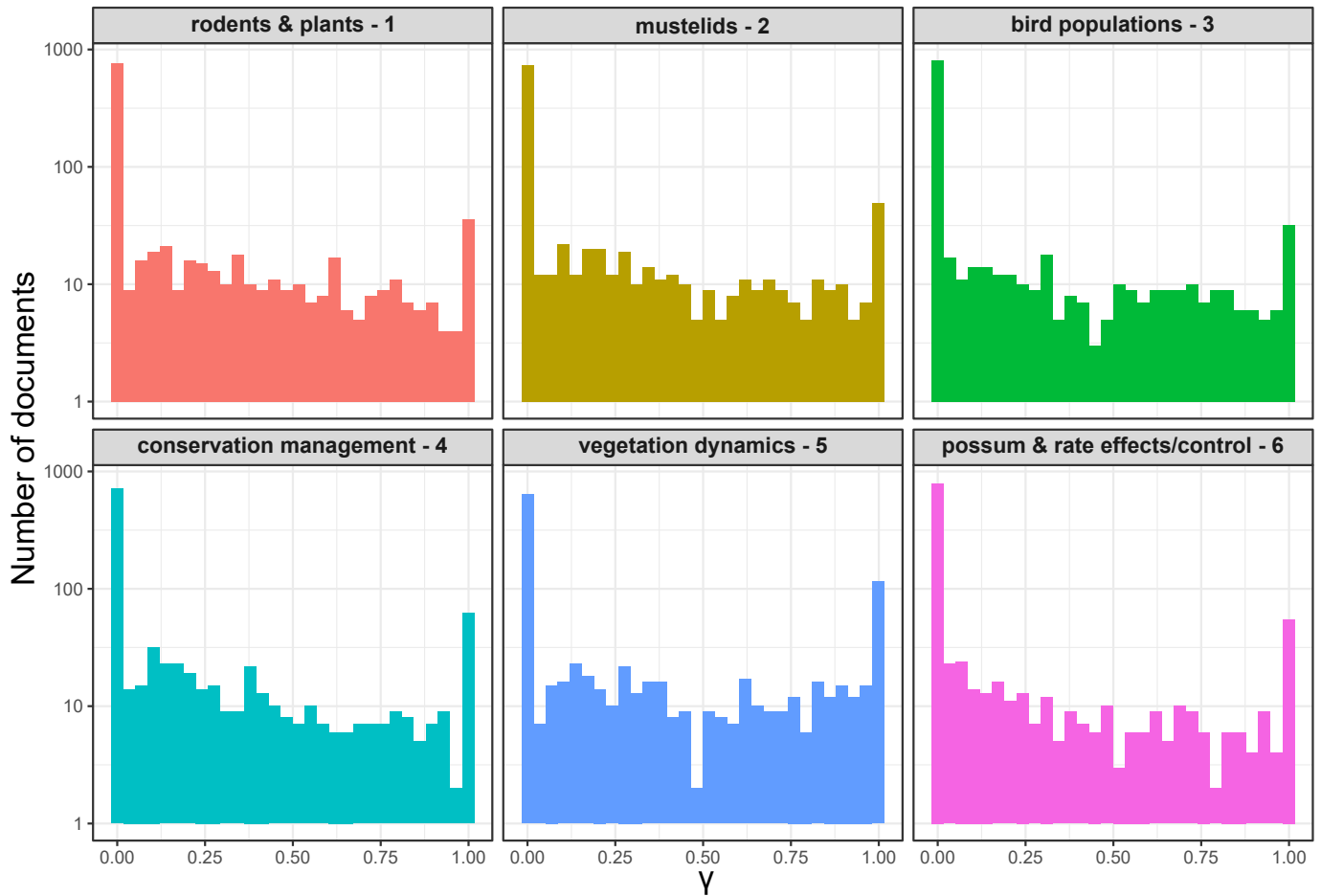


Supplementary Materials

Appendix S1. Network depiction of co-occurrence of words in titles in the *PESNZ* and *NZJE*. The width of the links is proportional to their frequency.



Appendix S2. Distribution of γ values (probability of membership) for each topic identified via Latent Dirichlet allocation (LDA). These distributions show that the model has reasonable classification power as most plots are strongly bimodal (either close to zero or close to one probability of membership); note the log-scaled y-axis.



Appendix S3. Stopwords used in the Latent Dirichlet allocation.

Stopwords are common words such as *about above, more, high* that are unlikely to identify latent topics; to facilitate topic identification we excluded stopwords from the list in the tidytext R library v 0.2.3 (Silge & Robinson 2016). We also excluded a suite of more specific terms that we found blurred topic identification: “new”, “zealand”, “new zealand”, “south”, “north”, “island”, “tiritiri matangi”, “central”, “otago”, “species”, “studies”, “study”, “ecology”, “ecological”, “region”, “site”, “westland”, “females”, “males”, “female”, “male”, “new society”, “c”, “ha”, “n”, “society”, “forest”, “forests”, “found”, “rate”, “data”, “effect”, “research”, and the digits 0 through 10.

Appendix S4. Rationalisation of species names and removal of place names in keywords (Fig. 8).

In the analysis of keywords, we rationalised species names where multiple terms were used for species that were unequivocally the same.

Trichosurus vulpecula → brushtail possum
 possum → brushtail possum
 brushtail possums → brushtail possum
 possums → brushtail possum
Mustela erminea → stoat
Rattus rattus → ship rat
Oryctolagus cuniculus → rabbits

Placenames removed: New Zealand, Canterbury, Stewart Island.