Quantifying the dynamics of nearly 100 years of dominance hierarchy research

Elizabeth A. Hobson[1,*]

[1]*Department of Biological Sciences, University of Cincinnati, Cincinnati, OH USA*

[*]*Corresponding author; e-mail: elizabeth.hobson@uc.edu*

Abstract

Dominance hierarchies have been studied for almost 100 years. A science of science approach can help provide high-level insight into how the dynamics of dominance hierarchy research have shifted or been maintained over this long timescale. To summarize these general patterns, I extracted publication metadata using a Google Scholar search of "dominance hierarchy", resulting in over 26,000 publications. I used text mining approaches to assess patterns in three areas: (1) general patterns in publication frequency and rate, (2) dynamics of term usage, and (3) term co-occurrence in publications across the history of the field. While the overall number of publications per decade continues to rise, the percent growth rate has fallen in recent years, demonstrating that although there is sustained interest in dominance hierarchies, the field is no longer experiencing the explosive growth it showed in earlier decades. Based on term co-occurrence networks and community structure, the different subfields of dominance hierarchy research were most strongly separated early in the field's history while modern research shows more evidence for cohesion and a lack of distinct term community boundaries. These methods provide a general view of the history of research on dominance hierarchies and can be applied to other fields or search terms to gain broad synthetic insight into patterns of interest, especially in fields with large bodies of literature.

Introduction

Competition is nearly ubiquitous in situations where resources are limited and contested. Because of this, conflict is inevitable in most social groups, leading to increased access to these resources for some individuals and decreased access for others. In many social species, this competition leads to the emergence of group dominance hierarchies, which can help make social life more structured and predictable and regulate overall conflict. In nearly 100 years of research on dominance, scientists have documented the presence of hierarchies that structure social conflict in a wide range of species [1; 2]: groups of ants, fish, lizards, geese, parrots, elephants, hyenas, pri-

28 mates, and many species in between form groups with detectable hierarchies, where individuals
29 within the groups can be assigned ranks. Much research has also established that these hierar-
30 chies also matter to individuals: higher-ranked individuals often benefit from improved health or ac-
31 cess to resources, more reproductive opportunities, more offspring, or greater longevity [3; 4; 5; 6].

32 Scientific progress itself is a social process, with new research continually building on the founda-
33 tions of previous work. By taking a science of science approach, we can synthesize the history
34 of a large and active field like dominance research. This approach complements more typical lit-
35 erature reviews with a "big data" perspective on publishing patterns and topics in the field. There
36 have been too many papers published on dominance hierarchies in too many subfields for it to be
37 feasible to read and synthesize the entire body of work. A traditional literature review may also
38 be unintentionally biased towards certain subtopics, study species, or subfields. The quantitative
39 approach used here allowed me to summarize a much bigger body of literature to identify general
40 patterns and provide an overall summary of how areas of focus and study have changed over the
41 course of nearly 100 years of research since the original publication describing "peck order" in
42 chickens [7]. However, it is important to note that what we gain from this high-level perspective
43 is balanced by the absence of important syntheses of knowledge that are part of good literature
44 reviews (see other papers in this special issue for this perspective).

45 Here, my goal is to quantify general patterns of interest in this long-running field of research and
46 identify patterns of interest, how the focus of research has shifted over time, and how themes
47 have waxed and waned across decades. Specifically, I focus on three main aspects: (1) general
48 patterns in overall publication frequency and rate, (2) dynamics of changes in term usage, and
49 (3) how terms co-occur in publications across the history of the field. I used a combination of
50 data scraping and text mining approaches to quantify these patterns. This approach provides a
51 broad review of the history of study of dominance hierarchies to better understand where and how
52 researchers have focused their scientific efforts.

Methods

Data collection and processing

55 I used a text mining approach to quantify long-term patterns. I determined when keywords were
56 highly used in titles based on publication decade and how these terms co-occurred with each
57 other in titles as a measure of the focus of research in each decade. I used the program "Publish
58 or Perish" [8; 9] to scrape Google Scholar for publications using the search term "dominance
59 hierarchy". The Google Scholar algorithm returns publications which match the search terms
60 anywhere in the searched documents (author, title, source, abstract, references, etc.) [9]. While
61 the full Google Scholar search algorithm is not publicly available, the algorithm generally works by
62 considering the full text of each document, the publication venue, the authors, and the recency and
63 frequency with which other papers have cited it to rank all publications and return the top 1000
64 results. I used Google Scholar as the main search engine because it has wider coverage and
65 returns more publications than many other searches (i.e. Web of Science, [10]). It also returns a
66 broader array of publication types including "grey literature" like theses, conference proceedings,
67 white papers, and preprints [11]. While using Google Scholar is beneficial for its wider reach,
68 there are some important limitations: it does not have strong quality control processes so data are
69 generally not as clean as the more limited output from other sources [10]. Results, including titles,

70 are also sometimes truncated if they are too long [9].

71 Using Publish or Perish, I collected the titles of the top 1000 Google Scholar results of publications
72 matching the search term for each year (with searches conducted from 2021-04-14 to 2021-04-
73 17). Any records missing a valid publication year, or with a publication year prior to 1920, were
74 excluded from the analyses to pin the start of analyses to the decade in which the foundational
75 paper for the field, by Schjelderup-Ebbe, was published in 1922 [7]. Analyses were pooled by
76 decade to better identify general patterns.

77 All titles in the database were cleaned prior to text analyses using the R packages tm and stringr [12;
78 13]: punctuation, numbers, and whitespaces were removed, and all was text converted to lower-
79 case. I removed all stopwords (extremely common words like a, is, the, etc.) using the snowball
80 stopword list in the R tidytext package [14]. These cleaned title terms formed the corpus. Prior
81 to full analyses, all terms in the corpus were also stemmed using the the tm package. Stemming
82 reduces instances of similar words (e.g., "ecological" and "ecology" both reduce to the stemmed
83 word "ecolog"). Limited further stemming efforts for commonly-used words were conducted by
84 hand (e.g., both "behavior" and "behaviour" were retained as separate terms following stemming
85 in the tm package; I collapsed these to "behavior").

86 ### *Quantifying publication trends over time*

87 To quantify how publications trends involving the search term "dominance hierarchy" have changed
88 over time, I counted the number of publications in the database for each decade. I then quantified
89 the lagged change in publication number by comparing each decade's total publications to the
90 previous decade. Finally, I quantified the decade over decade percent growth in publications by
91 dividing the lagged change in publication number by the number of publications in the previous
92 decade and converting to a percent.

93 ### *Estimating dynamics of term usage*

94 To estimate term use dynamics, I filtered the terms in the title corpus to retain only the stemwords
95 used in at least 25 publications in the database. This process excluded terms that were rarely
96 used but also helped exclude any terms that were reported by Google Scholar in a somewhat
97 corrupted format (for example, punctuation was sometimes introduced erroneously within words
98 in the search results, some words were split by truncation, etc.). With this cleaned corpus with
99 rarely-used terms excluded, I then determined whether each stemword was present or absent
100 in each document's title (this corrects for cases where a particular stemword was used multiple
101 times in a single title). I used presence data to then find the total number of publications per
102 decade where each stemword was present.

103 To determine how stemwords were used from decade to decade, I quantified the Shannon diversity
104 of stemword use in each decade, using the frequency with which stemwords were present in
105 titles combined with the total number of unique stemwords. I also quantified the number of novel
106 stemwords per decade, where "novelty" was defined as a unique stemword which appeared in a
107 title where it was not present in earlier decades in the corpus. Finally, using the entire history of
108 dominance hierarchy stemword corpus, I found the percent of total stemwords that were present
109 in titles in each decade.

To measure decade to decade similarity in stemword usage, I compared the presence of stemwords in each decade to the presence of stemwords in all other decades. I then found the number of terms that were present in both decades, the percent of words present in both decades, and the percent similarity of each decade compared to the decade it was most similar to. This analysis helps determine "hot spots" of term use similarity in documents published across different decades, where the more similar term use was, the more similar those decades would be when compared.

Determining differences in term co-occurrence

Knowing that a certain term was present in documents in a particular decade is helpful for assessing when terms emerge and are popular. However, even more insight can come from quantifying how terms co-occur in documents through the history of the field. To quantify term co-occurrence, I used the dataset including all stemwords used in at least 25 publications (described above).

To determine how stemword co-occurrences have changed over time, I found the network of co-occurring stemwords in each decade. Inspired by methods used to reconstruct the cultural evolution of a music genre [15], I used the R package igraph and fastgreedy community detection [16; 17], to identify communities of co-occurring stemwords by decade. I calculated the modularity of stemword co-occurrence networks in each decade, which measures how well the community-detection algorithm partitions a network into communities [16]. Finally, for each decade, I quantified how communities were interconnected or separated from each other [18]. I calculated a cohesion index [15] to represent the ratio of within-community edges compared to connections between stemwords assigned to different communities, using unweighted binary edges.

To track similarity in community composition across decades, I used Jaccard similarity to measure among community similarity in stemword composition between adjacent decades. Jaccard similarity measures overlap in membership between sets as the size of the intersection divided by the size of the union of two sets. In my case, I used it to find the proportion of stemwords found in a community in one decade compared to the stemword compositions of communities in the next decade. The similarity measure ranges from 0 (no overlap in terms) to 1 (exactly the same terms used in both communities). I used this approach because communities in each decade were assigned numerical codes, but "Community 1" in one decade is not necessarily comprised of a similar assortment of stemwords in the next decade - in other words, there was no consistent naming continuity in communities across decades. To detect similarity in community composition of stemwords, I then plotted results from each community in each decade as a flow diagram to visualize which communities were most similar to each other through time and how communities merged and split over time using the R package riverplot [19].

Results and Discussion

Quantifying publication trends over time

Analyzing publication trends can provide insight into overall interest in a field of study. Research on dominance hierarchies has resulted in an impressive number of publications in nearly 100 years of research, with over 26,000 publications in the scraped dataset which were returned from a keyword search for "dominance hierarchy". After data cleaning, 25,219 publications were retained for the

Figure 1: Publications by decade showing (a) total publications for each decade, (b) the number of publications in each decade compared to the previous decade, and (c) the percent growth in number of publications compared to the previous decade.
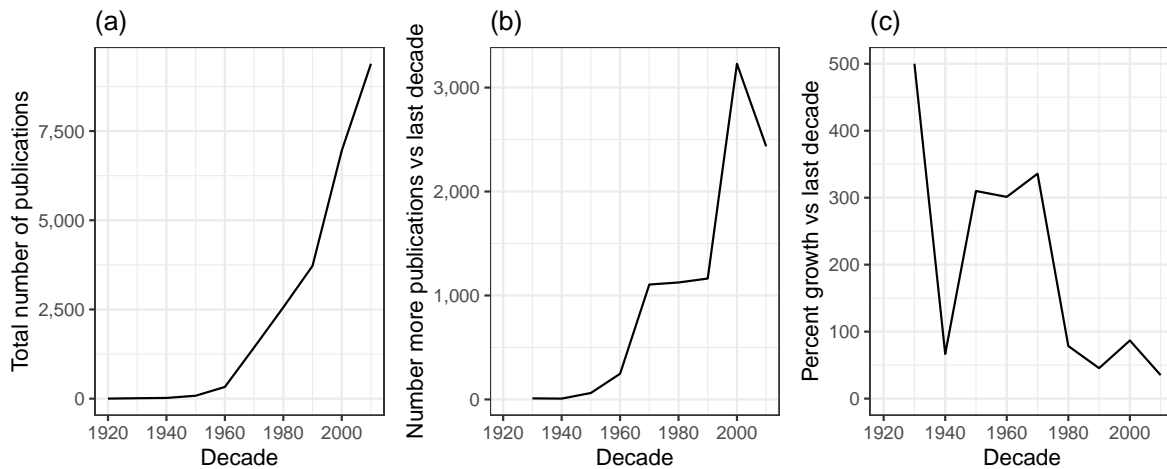

analyses. The highest number of publications per year to date occurred in 2017 (952 publications). The mean number of publications per year was 285.43 (median=166).

Figure 1 shows three views of publication trends by decade. While publication numbers have risen from decade to decade during the entire history of the field (Fig. 1a), it is also important to account for the overall increase in modern publication rates. Quantifying how the number of publications in one decade compares to the number of publications in the last decade helps somewhat normalize for this general increase in overall publications in modern science and helps better visualize changes in interest on a decade-by-decade scale. For example, the greatest increase in numbers of papers in a decade compared to the last decade occurred during the 2000's. 2010 was the first decade in which this explosive growth rate decreased: fewer additional papers were published in the 2010's compared to the increased number of publications when comparing the 2000's to the 1990's. The decade-over-decade percent growth rate provides different insight into publication trends, with the highest percent growth seen in the 1930's compared to the 1920's. This high growth rate indicates just how quickly the initial number of publications grew when compared to the "founding" of dominance hierarchy research in the 1920's. We see another peak in percent growth rates in the 1950's to the 1970's with decade over decade growth rates around 300%. These publication trends provide strong evidence for sustained interest in dominance hierarchy research despite nearly 100 years of study, but do indicate that in the most recent decades, the earlier explosive growth in numbers of publications has tapered off.

Estimating dynamics of term usage

While general patterns in the number of publications per decade provide evidence of sustained interest in dominance hierarchy research, how terms are used each decade provides insight into areas of focus for research efforts. I found that the overall diversity of stemword use per decade increased sharply up to 1950, then continued to increase at a slower rate until 1990 (Fig. 2a). The diversity of stemword use has been relatively stable from 1990 to 2020 and reflects the highest

5

Figure 2: Stemword usage in titles by decade showing (a) Shannon diversity in stemword use in each decade (with diversity calculated on total number of publications using each stemword per decade), (b) the number of novel stemwords introduced in each decade, and (c) the percent of stemwords in the entire corpus that were used in titles in each decade.
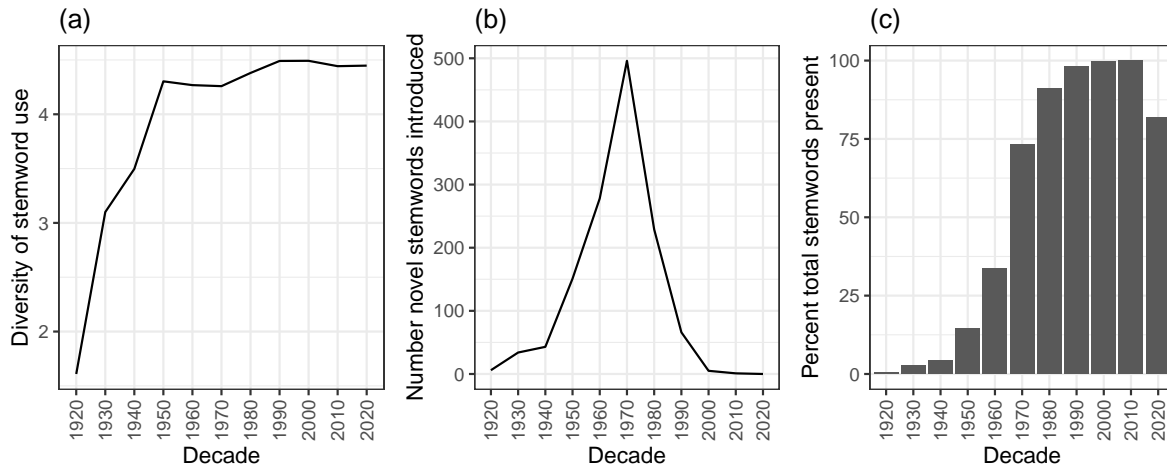
diversity period in the history of dominance hierarchy research. This high diversity period coincides with high overall numbers of publications during these decades, as seen in Figure 1. Interestingly, despite very few years of publications so far in the 2020's, publications from January 2020-April 2021 already share the high diversity of stemword use seen across the whole 2010 decade.

The number of novel stemwords introduced in publication titles peaked sharply in 1970 (Fig. 2b). On a decade by decade basis, Figure 2c shows how terms present across the entire history of dominance hierarchy research are used in publication titles: by 1970, nearly 75% of all stemwords in the corpus were in use.

Each decade can also be compared to other decades in the dataset to determine the levels of similarity in stemword use in titles over time (Fig. 3). The number of stemwords present in both decades peaks when comparing 1990, 2000, and 2010. Breaking these patterns down by the percent of terms present in both decades (compared to the number of terms present in either decade) shows an even larger hotspot of similarity that highlights how similar term use in 1980-2020 has been to other decades within that same time span. Comparing term use in each decade to the decade with which term use is most similar shows an even wider hotspot coinciding with all decades compared to 1990-2010; this hotspot also coincides with data in Figure 2c, showing that nearly 100% of terms used in the entire historical corpus were used in 1990-2010.

Determining differences in term co-occurrence

In addition to quantifying overall publication trends and the use of single terms in titles, the co-occurrence of terms in titles can provide even more insight into areas of focus for dominance hierarchy research. These term co-occurrences across decades is depicted in Fig. 4, where a community of stemwords in one decade that is very similar to the composition of stemwords in the next decade is connected with a thicker line. This plot also shows how stemword communities

Figure 3: Stemword usage in titles showing decadal similarity: (a) the raw number of terms present in both decades, (b) the percent of terms present in decades on the y-axis compared the terms present in decades on the x-axis, and (c) a measure of each decade's similarity scaled by maximum similarity to decades on the y-axis, with maximum similarity shown in red.
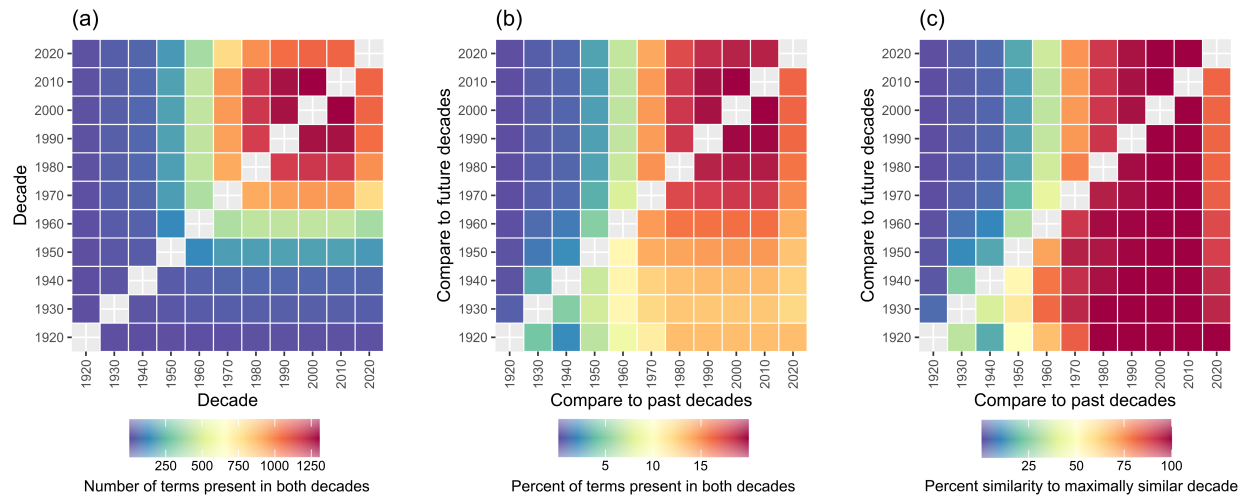
can split across decades as stemwords are assorted into new community configurations in the next decade. For example, Community 1 in 2000 is most similar to Community 1 in 2010, but shares some similarity with Communities 2 and 3, demonstrating that some stemwords contained in 2000's Community 1 were re-assorted and fell into separate communities in the next decade. Stemword community continuity was highest from 1990 to 2010, with an average maximal similarity of over 25% shared stemwords. Modern dominance research generally continues this trend of high continuity in stemword co-occurrence network structures. This pattern shifts back to multiple communities in the 2020's, but it is uncertain whether this pattern will persist as more publications are released in this decade.

While the composition of stemword communities has changed over time, the distinctness of borders between these communities has also changed. Figure 5 shows how stemword co-occurrence networks were much more modular early in dominance hierarchy research, but have decreased in modularity in recent years. The number of detected communities has also shifted from a small number of communities to a peak of 12 detected communities in 1950, followed by a gradual decline to fewer communities in modern dominance hierarchy research (it is uncertain if the rise in community number in the 2020's will persist as more papers are published in this decade so those results should be treated with caution). When I compared the percent of co-occurrences in stemwords in titles within stemwords in the same community compared to co-occurrences with stemwords assigned to different communities, I found that connections from one community across borders and to terms in another community has increased over the history of dominance hierarchy research to plateau in modern times at about 50%, indicating a balance between within-community and outside-community edges. This provides further insight into the factors contributing to the decrease in overall modularity as communities become more interconnected, as well as the reduction in the total number of detectable communities.

Figure 4: Title stemword community similarity over time: curved flow lines indicate similarity in stemword co-occurrence communities in one decade to the communities in the next decade (wider lines indicate higher similarity).
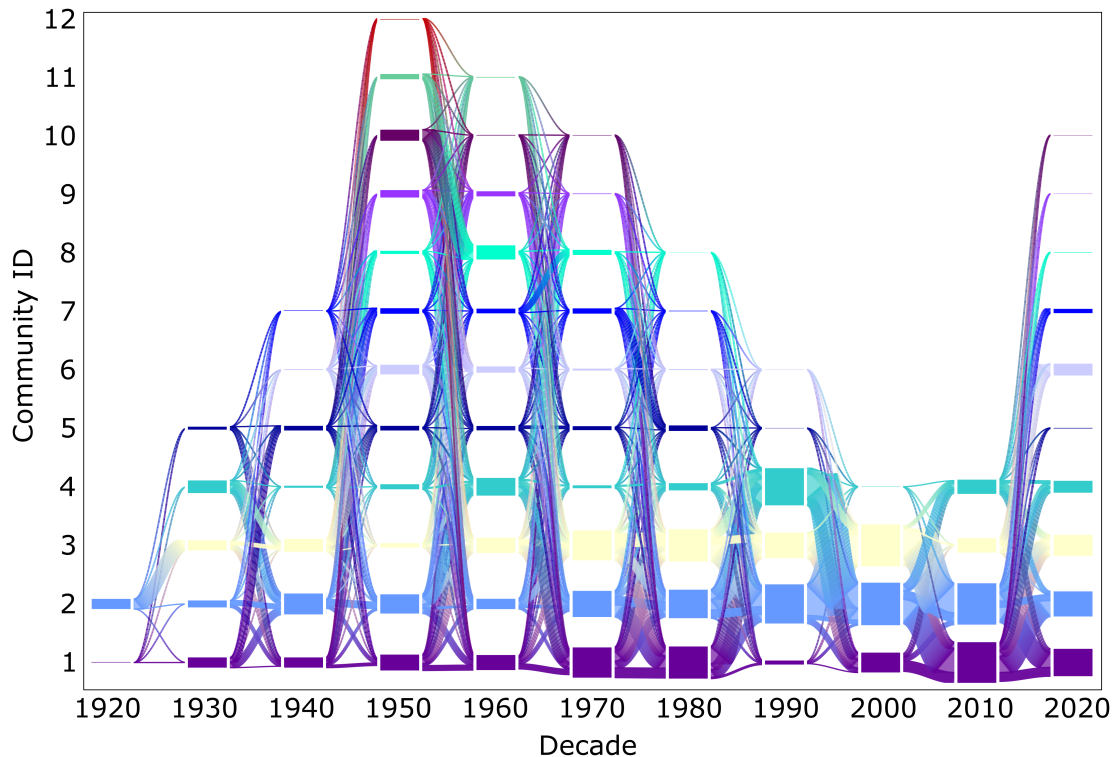

Conclusions

These analyses give insight into general publication trends, term use, and term co-occurrence in titles returned from a search for "dominance hierarchy" in publications across the nearly 100 years of dominance hierarchy research. From this, we can infer how investment in publishing dominance hierarchy research has changed, but also how connectivity between different subfields and topics has shifted.

This analysis of dominance hierarchy research is limited in that it only considers terms used in titles of publications. This analysis obviously cannot capture the complexity of how topics are treated in the full text of these publications, so cannot provide a detailed account of exactly how research trends or concepts have shifted over the history of dominance hierarchy research. However, this summary demonstrates that dominance research has had sustained interest over its long history and the evidence of especially connectivity does provide a high-level perspective on historical and modern trends in the science of hierarchies.

An open question amongst researchers working on dominance hierarchies is whether we have "solved" dominance. Informally, and depending on who you talk to at conferences, the question of dominance has been "solved" in the 1960's, the 1970's, the 1980's, or the 2000's. However, the sustained interest and investment in new publications demonstrated here, as well as the recent

Figure 5: Stemword co-occurrence network summaries by decade: (a) network modularity over time, (b) number of communities detected in each decade, and (c) percent of edges connecting a stemword node in one community to a stemword node in a different community.
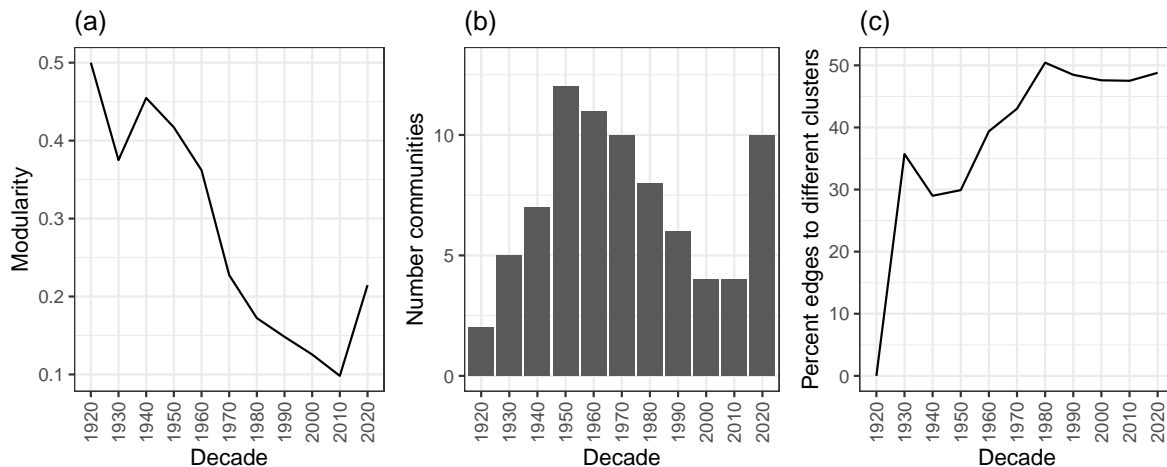
shift in the apparent focus of the field along with greater cohesion and cross-community connec-
tions, suggests that there is still much interest in all that we still have to learn about dominance
hierarchies. Recent dominance research may be moving to a new stage of research focus, par-
ticularly if we have solved some of the more basic hierarchy questions. In particular, new genetic
methods (e.g., [20], this issue), computational approaches [21], and a focus on the information
contained in both networks of aggression and rank within social groups [21; 22] provide many new
avenues for novel insight into animal sociality. Both theoretical work (e.g., [23; 24] and empiri-
cal work [21] have also recently suggested that rank acquisition can be remarkably sensitive to
stochastic events. Other empirical work has shown that "rule-breaking" via coalition formation can
cause disruptions to expected rank inheritance patterns [25], and that these dynastic changes can
gain momentum and persist despite the lack of underlying characteristics or quality to differentiate
these individuals from less-successful ones in the group. The integration of these new tools as well
as new, more complex ways of studying the decisions animals make about who, when, and how
they fight each other, provide a strong foundation for the next 100 years of dominance hierarchy
research.

References

[1] Shizuka, D. & McDonald, D. B. The network motif architecture of dominance hierarchies. *Journal of The Royal Society Interface* **12**, 20150080–20150080 (2015). URL `http://rsif.royalsocietypublishing.org/cgi/doi/10.1098/rsif.2015.0080`.

[2] Shizuka, D. & McDonald, D. B. A social network perspective on measurements of dominance hierarchies. *Animal Behaviour* **83**, 925–934 (2012).

[3] Sapolsky, R. M. Social Status and Health in Humans and Other Animals. *Annual Review of Anthropology* **33**, 393–418 (2004). URL `http://www.annualreviews.org/doi/10.1146/annurev.anthro.33.070203.144000`.

[4] Archie, E. A., Altmann, J. & Alberts, S. C. Social status predicts wound healing in wild baboons. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 9017–9022 (2012). URL `www.pnas.org/cgi/doi/10.1073/pnas.1206391109`.

[5] MacCormick, H. A. *et al.* Male and female aggression: lessons from sex, rank, age, and injury in olive baboons. *Behavioral Ecology* **23**, 684–691 (2012). URL `https://academic.oup.com/beheco/article/23/3/684/225015`.

[6] Sapolsky, R. M. The influence of social hierarchy on primate health. *Science (New York, N.Y.)* **308**, 648–52 (2005). URL `http://www.ncbi.nlm.nih.gov/pubmed/15860617`.

[7] Schjelderup-Ebbe, T. Beiträge zur Sozialpsychologie des Haushuhns [Observation on the social psychology of domestic fowls]. *Zeitschrift für Psychologie* **88**, 225–252 (1922). URL `https://psycnet.apa.org/record/1926-07009-001`.

[8] Tarma Software Research Ltd. Harzing's Publish or Perish (version 7.31.3306.7768) (2021).

[9] Harzing, A.-W. *The Publish or Perish Book: Your guide to effective and responsible citation analysis* (Tarma Software Research, Melbourne, Australia, 2013), html edition edn. URL `https://harzing.com/popbook/`.

[10] Harzing, A.-W. & Alakangas, S. Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics 2015 106:2* **106**, 787–804 (2015). URL `https://link.springer.com/article/10.1007/s11192-015-1798-9`.

[11] Haddaway, N. R., Collins, A. M., Coughlin, D. & Kirk, S. The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. *PLOS ONE* **10**, e0138237 (2015). URL `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0138237`.

[12] Wickham, H. *stringr: Simple, Consistent Wrappers for Common String Operations* (2019). URL `https://CRAN.R-project.org/package=stringr`. R package version 1.4.0.

[13] Feinerer, I., Hornik, K. & Meyer, D. Text mining infrastructure in r. *Journal of Statistical Software* **25**, 1–54 (2008). URL `http://www.jstatsoft.org/v25/i05/`.

[14] Silge, J. & Robinson, D. tidytext: Text mining and analysis using tidy data principles in r. *JOSS* **1** (2016). URL `http://dx.doi.org/10.21105/joss.00037`.

[15] Youngblood, M., Baraghith, K. & Savage, P. E. Phylogenetic reconstruction of the cultural evolution of electronic music via dynamic community detection (1975-1999) (2020). URL `https://arxiv.org/abs/2011.02460v2`.

[16] Clauset, A., Newman, M. E. J. & Moore, C. Finding community structure in very large networks. *Physical Review E* **70**, 066111 (2004). URL `https://journals.aps.org/pre/abstract/10.1103/PhysRevE.70.066111`.

[17] Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **Complex Systems**, 1695 (2006). URL `http://igraph.org`.

[18] McDonald, D. B. & Hobson, E. A. Edge weight variance: population genetic metrics for social network analysis. *Animal Behaviour* **136**, 239–250 (2018).

[19] Weiner, J. *riverplot: Sankey or Ribbon Plots* (2017). URL `https://CRAN.R-project.org/package=riverplot`. R package version 0.6.

[20] Anderson, J. A. *et al.* Distinct gene regulatory signatures of dominance rank and social bond strength in wild baboons. *bioRxiv* 2021.05.31.446340 (2021). URL `https://www.biorxiv.org/content/10.1101/2021.05.31.446340v1https://www.biorxiv.org/content/10.1101/2021.05.31.446340v1.abstract`.

[21] Hobson, E. A., Mønster, D. & DeDeo, S. Aggression heuristics underlie animal dominance hierarchies and provide evidence of group-level social information. *Proceedings of the National Academy of Sciences* **118** (2021). URL `https://www.pnas.org/content/118/10/e2022912118.short`.

[22] Hobson, E. A. Differences in social information are critical to understanding aggressive behavior in animal dominance hierarchies. *Current Opinion in Psychology* **33**, 209–215 (2020).

[23] Kawakatsu, M., Chodrow, P. S., Eikmeier, N. & Larremore, D. B. Emergence of Hierarchy in Networked Endorsement Dynamics. *PNAS* **118** (2021). URL `https://www.pnas.org/content/118/16/e2015188118https://www.pnas.org/content/118/16/e2015188118.abstract`.

[24] DeDeo, S. & Hobson, E. A. From equality to hierarchy (2021). URL `https://doi.org/10.1073/pnas.2106186118`.

[25] Strauss, E. D. & Holekamp, K. E. Social alliances improve rank and fitness in convention-based societies. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 8919–8924 (2019). URL `http://www.ncbi.nlm.nih.gov/pubmed/30858321http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6500164`.