How do we copy and paste? The semantic drift of quotations in blogspace

SÉBASTIEN LERIQUE and CAMILLE ROTH

1. INTRODUCTION

The understanding of the mechanisms behind cultural similarity and diversity has led to a sizeable literature in the recent past, spanning over a vast area of research fields ranging from cultural anthropology to social network analysis and complex systems modelling, all diversely labelled as studies on "opinion dynamics", "cultural evolution", or "information diffusion", for the most part. Broadly, this type of research program investigates phenomena pertaining to both cognitive science and social science, and aims to understand the processing, transmission, and evolution of information both at the individual and social levels.

Several theories have been proposed and debated within these research fields, mixing social and individual cognition, notably from the side of cultural anthropology. First, the debate around the "memetic" program initiated by Dawkins [1976], for which the collection of works by Aunger [2000] provides a solid overview. A second field has focused on the development of evolutionary models of norms (see for instance Ehrlich and Levin 2005), following the seminal work of Boyd and Richerson [1985].

A third approach, "cultural epidemiology", initiated by Sperber [1996] recently attracted significant attention. Works such as Atran [2003] argue that this approach is anthropologically better suited than memetics, and the main issues in this debate are further detailed by Kuper [2000] and Bloch [2000]. The cultural epidemiology program focuses on the notion of *representation*, linking the cognitive science concept of *mental representation* to the concept of *public representation*, the latter being the former's counterpart outside the brain (i.e. in cultural artefacts: texts, utterances, etc.).[1] The key point here is that representations are not being replicated through a high-fidelity copy process, but are being interpreted and produced anew, and are thus greatly subject to change. Cultural epidemiology postulates this type of conceptual evolution and suggests that it can be appraised through the notion of "cultural attractor", seen as the attraction domain of an underlying socio-semantic dynamical system. Despite some recent modelling attempts (for instance Claidière and Sperber [2007]), the development of quantitative measurements focused on the key notion of cultural attractor has remained a relatively hard task and, to our knowledge, this hypothesis has not yet been empirically analysed.

However, the last decade has witnessed an avalanche of observable *in vivo* data in the form of online interactions. While they are not records of "physical" inter-individual interactions (in the sense of "real life" interactions), these productions and information trails still constitute a wealth of observations on the dynamics of public – albeit online – representations. Given the already significant – and rapidly growing – importance of our online interactions, these records can dramatically improve the prospects of empirical study of the individual-level processes of cultural evolution.

We aim to empirically describe the transformation of a specific type of public representation, by focusing on the possible alterations introduced by individuals when newly producing a representation. To deal with robust and simple cultural representations, we paid attention to the evolution of quotations. While these verbatim public representations should in theory not suffer any alterations as they are produced anew (as opposed to more elaborate expressions and opinions, not identified as quoted utterances), empirical observation shows that they are in fact quite often transformed. We will in particular exhibit a non-trivial process by which individual words in quotations are replaced. We will uncover some of the semantic and structural characteristics of these words and the substitutions they undergo. More generally, we contend that using this type of data is equivalent to a large-scale psycholinguistic experiment and at the same time constitutes the first step towards building empirically realistic models of cultural evolution.

The next section (Sec 2) describes the state-of-the-art on this matter. In Sec. 3, we detail the empirical protocol and the various assumptions that were made in order to deal with the available empirical material. Section 4 describes the significant psycholinguistic cues and biases observed during *in vivo* quotation reformulation, followed by a discussion and general guidelines for further work in Sec. 6.

2. RELATED WORK

The relevant literature on *public representation dynamics* features two main streams. On one hand, we find studies of the macroscopic *social diffusion* of public representations, describing for instance the propagation of cultural artefacts across social networks such as blogspace [Gruhl et al. 2004], the characteristic times and diffusion cycles both within these social networks and with respect to the topical dynamics of news media [Leskovec et al. 2009a], or the reciprocal influence between the social network topology and the distribution of issues [Cointet and Roth 2009]. These studies are relatively independent from anthropology and cognition and are at the interface between data mining, complex systems and quantitative sociology (first and foremost social network analysis). Without necessarily relying on specific social science theories, this research stream is of interest for its use of large social media corpora in studying cultural dynamics.

On the other hand, the study of the *transformation* of public representations has emerged only recently. For one, models involving evolution and representations to study the notion of "cultural attractor" have appeared only a few years ago [Claidière and Sperber 2007]. Among the empirical approaches on the mutation of representations, some of the most relevant studies to date consist in a series of papers investigating *quotation* transformations in a large corpus of US blog posts, initially collected and studied by Leskovec et al. [2009a] and further analyzed by Simmons et al. [2011] and Omodei et al. [2012]. They show several types of regularities and

[1] Sperber [1996] emphasizes this distinction in his seminal work:

> A representation may exist inside its user: it is then a *mental representation*, such as a memory, a belief, or an intention. The producer and the user of a mental representation are one and the same person. A representation may also exist in the environment of its user, as is the case, for instance, of the text you are presently reading: it is then a *public representation*.

propose diffusion-transformation models of the evolution of quotations, which may nonetheless appear to be relatively simplistic from a cognitive viewpoint. One of the main conclusions of these works is that even for quotations, a type of public representation that should be among the most stable, it is still possible to observe and measure significant transformations. However, these studies address transformations by focusing on the properties of the source of the quotation (e.g. news outlet v. blog), or the surrounding public space (e.g. quotation frequency in the corpus), rather than the very cognitive-level features which may determine or, at least, influence these transformations.

At this level, we have to turn to the broader psycholinguistic literature which provides one of the main cognitive foundations for public representation evolution by studying the influence of word features on the ease of recall. This field is well developed and details the impact that classical psycholinguistic variables such as word frequency (see Yonelinas [2002] for a review), age-of-acquisition [Zevin and Seidenberg 2002], number of phonemes or number of syllables (see for instance Rey et al. [1998] and Nickels and Howard [2004]), have in this type of task.

Less classical linguistic variables, based on the study of semantic network properties, have recently appeared as an empirical investigation field, after having been heavily discussed around the notion of connectionism and its normative processual models (see for instance Collins and Loftus [1975]). Let us mention two interesting and recent studies on that matter, which demonstrate in a strictly *in vitro* framework and at the vocabulary level that word properties computed on a word network are important factors for the cognitive processes and reproduction of those words. First, Griffiths et al. [2007] analyse a task where patients are asked to name the first word which comes to their mind when they are presented with a random letter from the Latin alphabet. The authors show that there exists a link between the ease of recall of words and one of their semantic features, namely their authority position (PageRank) in a language-wide semantic network built from external word association data. A second psycholinguistic study by Chan and Vitevitch [2010] shows, in a picture-naming task, that words come quicker when they have a higher clustering coefficient in an underlying phonological network (which, again, is defined from external phonological data).

[add *https://www.sciencedirect.com/science/article/pii/S0893608012000330, http://research.clps.brown.edu/austerweil/pdfs/papers/randomWalkNips2012.pdf, http://link.springer.com/article/10.3758/s13421-013-0312-y* ***to that picture]***

On the whole, the current psycholinguistic state-of-the-art seems to hint towards two antagonistic types of results. On one hand, part of the literature tends to show that recall is easier for the least "awkward" words; those whose age of acquisition is earlier, length is smaller, semantic network position is more central – this is particularly true in tasks where participants are asked to form spontaneous associations or utter a word in response to a given signal [Citation needed]. On the other hand, when the task consists in remembering a specific list of items, "awkward" words are actually more easily remembered, possibly as they are more informative and plausibly more discernible [Citation needed]. The jury is still out as to whether reformulation alteration, i.e. spontaneous replacement of words when asked to repeat a given utterance, is rather of the former or latter sort. Our paper additionally sheds light on this debate. ***[Show the link with epidemiology, and fitness: oddness is a kind of fitness]***

Stepping back, we observe a gap between, on one side, macro-level empirical studies of the diffusion dynamics in a social system and, on the other side, studies focused on micro-level transformations of representations — these latter studies being either strongly normative, or whose results are difficult to articulate with realistic cultural epidemiology models.

3. PROTOCOL

In order to start bridging this gap, we set out to *empirically* study public representation transformations at the microscopic level, aiming to stay compatible with macroscopic-level studies of these public representations. Quotations appeared to be a perfect candidate as public representations. First, they are usually cleanly delimited by quotation marks (and often with HTML markup in web pages), which greatly facilitates their detection in text corpora. Second, they stem from a unique "original" version, and could ideally be traceable back to that version. Third, and most importantly, their duplication should *a priori* be highly faithful, apart from cases of cropping: not only should transformations be of moderate magnitude, but when specific words are not perfectly duplicated, it is safe to assume that the variation is due to involuntary cognitive bias — as writers may expect any casual reader to easily verify, and thus criticize, the fidelity to the original quotation. Quotation evolution is therefore a perfect environment to measure cognition-induced transformations and relate those findings to macroscopic social dynamics.

3.1 Dataset

We relied on a reliable quotation dataset collected by Leskovec et al. [2009a], large enough to lend itself to statistical analysis. This dataset consists of the daily crawling of news stories and blog posts from around a million online sources, with an approximate publication rate of 900k texts per day, over a nine-month period of time (from August 2008 to April 2009) [Leskovec et al. 2009b].[2] Quotations were then automatically extracted from this corpus: each quotation is a more or less faithful excerpt of an utterance (oral or written) by the quoted person. Quotations were then gathered in a graph and connected according to their similarity: either because they differ by very few words (in that case, no more than one word) or because they share a certain sequence of words (in that case, at least ten consecutive words). A community detection algorithm was applied to that quotation graph to detect aggregates of tightly connected, i.e. sufficiently similar, groups of quotations (see Leskovec et al. [2009a] for more detail). This analysis yielded the final data we had access to, with a total of about 70,000 sets of quotations; each of these sets allegedly contains all variations of a same parent utterance, along with their respective publication URLs and timestamps.

3.2 Word-level measures

To keep the analysis palatable, we restricted the analysis to quotation transformations which consisted in the *substitution* of a word by another word (and only those cases). To quantify those substitutions, we decided to associate a number of features to each word, the variation of which we can statistically study. The following sections detail the features we used.

3.2.1 *Standard psycholinguistic indices.* We first introduce some of the most classical psycholinguistic measures on words:

[Add some bibliography about those features' known effects]

[2] Unfortunately, the original article [Leskovec et al. 2009a] does not provide additional details on the source selection methodology.

—[*Add word frequency as a feature, and maybe link it to Kolmogorov's 5/3 law in turbulences*]

—**Age of Acquisition**: the average age at which words are learned, obtained from Kuperman et al. [2012],

—The average **Number of Phonemes** for all pronunciations of a word, obtained from the Carnegie Mellon University Pronouncing Dictionary [Weide 1998],[3]

—The average **Number of Syllables** for all pronunciations of a word, also obtained from the CMU Pronouncing Dictionary.

We also considered grammatical types within quotations by detection of *Part-of-Speech* (POS) categories, using the Penn TreeBank Project typology [Santorini 1990] and thereby distinguishing verbs, nouns, adjectives and adverbs. The results were however extremely similar across the various categories, exhibiting no specific effect of words belonging to different POS categories.

3.2.2 *Network-based measures.* Additional to classical psycholinguistic measures, we also considered more recently studied variables based on semantic network properties. We relied on two types of network-based data: curated synonym associations provided by WordNet [WordNet 2010], and "free association" norms collected by Nelson et al. [2004].

The WordNet (WN) database gathers more than 117,000 concepts and about 147,000 words attached to these concepts. WordNet entries are disambiguated words, called "lemmas", which are grouped into concepts called "synsets": all the lemmas in a synset are synonyms for the same single meaning. Note that synsets also encode grammatical category, so the choice of a synset for a word (thus creating a lemma) immediately determines the grammatical category of that word. Non-disambiguated words make the nodes of the semantic network we build (i.e. homograph lemmas are aggregated into a single node), and a node's neighbours are synonymous lemmas belonging to the same synset(s) as the lemma(s) represented by that node. In other words, synsets induce cliques in the semantic network, and a word with several meanings belongs to as many cliques as it has meanings. Links connecting two synonymous lemmas are assigned a weight corresponding to the number of times these lemmas are mentioned by WordNet as synonyms, i.e. in the same synset. [*We might spare ourselves that last sentence since we drop that aspect further down.*]

Free Association (FA) norms record the words that come to mind when people are presented with a given cue (that is the "free association" task). As Nelson et al. explain,

> free association response probabilities index the likelihood that one word can cue another word to come to mind with minimal contextual constraints in effect. [Nelson et al. 2004]

Following Griffiths et al. [2007], we consider the directed weighed network formed by the association norms, that is the network where words are nodes and edges are directed from cue to associated word, with a weight equal to the probability of that target word being produced when this particular cue was presented.

While the original WN network is a weighted, undirected network, the FA network is a directed network where weights have a different meaning. Both semantic networks describe synonymy relationships corresponding to qualitatively distinct forms of association: either from a linguistic viewpoint (WN) or from the perspective of actual information retrieval by humans (FA). In order to work in a unified framework, we shall consider links as an information on the existence of some sort of association between two words and transform both networks into undirected, unweighed networks. [*Roll back to directed FA, and explain that it's a link between in-vivo spontaneous substitution and in-vitro forced substitution. We could even add a more precise analysis using FA data directly (instead of features on the FA network), or keep that for the processual model.*]

We introduce three standard network-based measures to be used on both the WN and FA networks:

—**Distance** between two words, a classical dyadic network measure corresponding to the shortest path length between two nodes. This measure may be used as a proxy of the remoteness between two given words, which is likely to influence the likelihood of a word being a relevant substitution candidate for another word.

—**Centrality** k, measured by the number of neighbours of a given node, which can be interpreted as a proxy for the polysemy of words. For the directed network FA we also define the incoming centrality k' (number of neighbours pointing *to* that node). [*Rephrase once we rolled back FA to directed*] Note that in the present case, there is a quasi-perfect correlation between node degree and node *PageRank* [Page et al. 1999], which had already been used in Griffiths et al. [2007] and may be interpreted as a generalized and recursive measure of word polysemy: central nodes in the PageRank sense are those which are linked to many nodes themselves linked to many nodes, and so on recursively.

—**Clustering coefficient** c, which measures the extent to which a node belongs to a local aggregate of tightly connected nodes, and defined as the ratio between the number of actual v. possible edges between a node's neighbours [Watts and Strogatz 1998]. In the WN network, a word is likely to have a higher clustering coefficient whenever the synsets to which it belongs are themselves semantically linked. This measure had been used by Chan and Vitevitch [2010].

Betweenness coefficient, another measure of node centrality describing the extent to which a node tends to connect otherwise remote areas of the network [Freeman 1977]. More technically, it corresponds to the normalized number of shortest paths connecting dyads which pass through that node; the higher the coefficient, the more important that node is in ensuring the connectedness of the rest of the network. Correlation between FA-BC and FA-PR is 0.74; between WN-BC and WN-PR it is 0.83. [*We dropped this, right?*]

3.2.3 *Variable correlations.* An important question here relates to the various possible correlations between all the variables we consider.

The age of acquisition is a key variable, primarily as a usual suspect in psycholinguistic studies, and also appears to be generally correlated to many of the other variables. This relates to an ongoing debate suggesting that age of acquisition encodes a variety of phenomena which are difficult to disentangle from more specific phenomena, which could be captured by more independent variables such as [Citation needed]. [*What do we conclude of that? Nuance results with AoA?*] Indeed, the minimal linear correlation of aoa (-0.31) with other variables has a higher magnitude than the maximal linear correlation of other variables with each other (-0.26). [*how do we get to that conclusion?*]

[3]The CMU Pronouncing Dictionary is included in the NTLK package [Bird et al. 2009], the natural language processing toolkit we used for the analysis.

Additionally, number of phonemes and number of syllables exhibit a strong linear correlation (0.85). The analysis showed a better prediction effect of number of phonemes over number of syllables, which is consistent with Nickels and Howard [2004], and we therefore chose to focus the presented results on the former only. ***[Rather, graphs are much nicer with number of phonemes than number of syllables: clearer tendencies, more data (there's holes with number of syllables). But not a prediction (prediction of what?)]***

Network properties, finally, appear to be relatively weakly correlated with each other, across one network or between both networks (i.e. nodes with high index values are generally distinct between FA and WN). We chose to focus the results on FA as this network presents a direct psychological interpretation (that human copy-pasting may or may not show the same type of trend as human free association), noting that prediction results are matched against WN as well. ***[comment peut-on dire que les nœuds importants dans FA ne sont pas nécessairement importants dans WN mais que, pourtant, la forme de la prédiction est la même? cela me semble illogique en première lecture] [Check those assertions]***

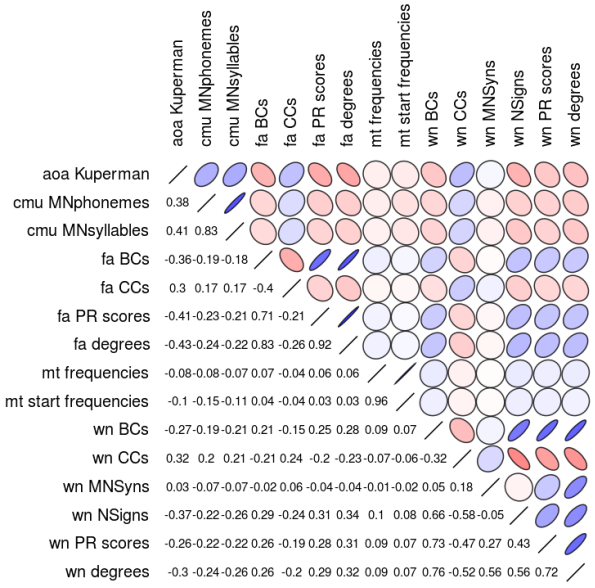


Fig. 1: Variable correlations ***[rebuild with the variables we talk about and keep]***

3.3 Detection of substitutions

The data we use presents an additional challenge: each set of quotations bears no explicit information either about the authoritative original quotation, or about the source quotation(s) each author inspired himself from when creating a new post and reproducing (possibly altering) those sources. Substitutions are therefore not explicitly encoded in the dataset.

We face a reverse engineering problem where, given all quotations and their occurrence timestamps, we must estimate which was the originating quotation for each instance of each quotation. We therefore model the underlying quotation selection process by making a few additional assumptions which let us define substitutions from the available data.

The main question in this problem concerns the way we consider later occurrences of a quotation which, when it first appeared, was identified to be an alteration from an original quotation. Let us give an example: say the quotation "These accusations are false and absurd" (q_a) appears in a blog on January 18, and the slightly different quotation "These accusations are false and incoherent" (q_b) appears in another blog on the 19th, the 20th, and the 21st of January. If q_a was sufficiently prominent when q_b first appeared, we can safely assume that the author of q_b on the 19th based himself on q_a. But what about the following occurrences of q_b? Should we consider them to be substitutions based on q_a (i.e. re-creations of q_b by a new instance of the substitution process that brought from q_a to q_b in the first place) or reproductions of the first occurrence of q_b?

To settle this question we model the process as follows: we assume that when a quotation q appears at time t, if it is not original (i.e. if not stemming from a source external to the dataset, e.g. initiating a new set of quotations), then it is based solely on the most frequent quotation q_{max} in the preceding period of time $[t - \Delta t; t]$. The length Δt of that period of time is fixed as a fraction of the total duration of the set of quotations; we took one fifth in the implementation (i.e. one month for a set of quotations spanning five months, or one day for a set spanning only five days; this takes the dynamism of the set of quotations into account). If q differs from q_{max} by only a word, it is counted as a substitution from q_{max} to q. In any other case, i.e. if q and q_{max} are the same or if q and q_{max} are different in other ways, the occurrence of q is not considered to be an instance of substitution and is discarded.

An example of this detection in a situation akin to the one described above can be seen on figure 0**??**: q_a and q_b differ by only a word, and q_a appears first and stays the most frequent in the beginning. This is why the occurrences of q_b at t_1 and t_2 are detected as substitutions stemming from q_a. After a time the situation is reversed: q_b becomes more frequent than q_a. This entails that occurrences of q_b are seen as reproductions of itself (t_3), and that occurrences of q_a are detected as substitutions stemming from q_b (t_4), i.e. *re-creations* of q_a.

[Add figure describing the chosen model]

The assumptions embedded in this model are only a subset of a wider set of possibilities, each leading to alternative models. We identified and implemented five other such models, and they all yielded essentially the same results. These models differ in the definition of the source quote from which new occurrences stem, essentially modifying the balance between reproduction of previous occurrences of a quote and re-creation of itself by a new substitution instance. For example, the time-windows considered can have different lengths, can include all occurrences from the beginning of the set of quotations, or can have fixed positions. The source quotation of a potential substitution can be chosen among those time-windows, or otherwise (e.g. among *all* quotations having appeared before t and differing by a word from the arrival quotation; this detection process would include all possible substitutions detected by other models, but would also include many false positives).

3.4 Characterization of substitutions

We then measure the alterations introduced when authors reproduce quotations by comparing the relative features of substituted and substituting words. Note that since we only consider substitutions, we measure the features of an alteration *knowing that there has been a alteration*, and we do not take invariant quotations into account. Indeed, in the first case we know there has been a human

reformulation, whereas in the second case it is impossible to know whether there has been perfect human reformulation or simply digital copy-pasting of a source ("CTRL-C/CTRL-V").

We build two main observables for each word feature. First, we measure the variation of word feature over a substitution, looking at the variation of a given feature between arrival and start words. Second, we measure the susceptibility for words to be the source of a substitution, knowing that there has been a variation, in order to show which semantic features are the most likely to attract a substitution.

3.4.1 *Alteration.* For a given feature $\mathfrak{F}$, we measure how a word w's feature varies as w is substituted by $w^\star$. Let us denote this quantity:

$$\Delta(w, w^\star) \stackrel{\text{def}}{=} \mathfrak{F}(w^\star) - \mathfrak{F}(w)$$

Averaging this value over all start words with a given feature value f yields the mean variation for that feature value f. [4] This quantity can be written:

$$\langle \Delta(w, w^\star) \rangle_f = \langle \mathfrak{F}(w^\star) - \mathfrak{F}(w) \rangle_{\{(w,w^\star)|\mathfrak{F}(w)=f\}}$$

We introduce a null hypothesis $\mathcal{H}_0$ to compare the actual variation of a word's feature to its expected variation, assuming the arrival word $w^\star_{\mathcal{H}_0}$ had been chosen randomly from the pool of free association words. The corresponding average quantity over all start words may be written: [5]

$$\left\langle \Delta(w, w^\star_{\mathcal{H}_0}) \right\rangle_f = \left\langle \mathfrak{F}(w^\star_{\mathcal{H}_0}) - \mathfrak{F}(w) \right\rangle_{\left\{(w,w^\star_{\mathcal{H}_0})|\mathfrak{F}(w)=f\right\}}$$

We also considered an alternative null hypothesis, denoted $\mathcal{H}_{00}$, where the arrival word is chosen randomly *among immediate synonyms of the start word* (neighbours in the WN network ***[We could also do that with neighbours in directed or undirected FA]***), i.e. an arrival word chosen among semantically plausible though still random words. [6]

Using this method we obtain the mean variation of feature for each start feature value, and can compare the variations to a situation where arrival words are chosen randomly. This gives us a fine-grained view of how word features evolve upon substitution.

3.4.2 *Susceptibility.* Furthermore, the protocol lets us compute substitution *susceptibilities* for each feature value f. We say that a word is *substitutable* if it appears in a quote which undergoes a substitution, whether that substitution operates on the considered word or on another. Word substitution susceptibility (denoted $\mathfrak{S}_\mathfrak{F}(w)$) is computed as the ratio of the number of times $n_s(w)$ a word is substituted to the number of times $n_p(w)$ that word appears in a substitutable position. We have:

$$\mathfrak{S}_\mathfrak{F}(w) \stackrel{\text{def}}{=} \frac{n_s(w)}{n_p(w)}$$

Now averaging over all words having a given feature value f, we obtain the mean susceptibility for the feature value f:

$$\langle \mathfrak{S}_\mathfrak{F} \rangle_f = \left\langle \frac{n_s(w)}{n_p(w)} \right\rangle_{\{w|\mathfrak{F}(w)=f\}}$$

This measure focuses on the selection of start word instead of the selection of the arrival word. Indeed, the features have an effect not only on the choice of a new word when a substitution takes place, but also at the preceding moment when it is not yet known which word in the quotation – if any – will be substituted.

4. RESULTS

We may see that $\mathcal{H}_0$ and $\mathcal{H}_{00}$ are slightly translated yet not qualitatively different.

4.1 Location of arrival words

Where in the network are arrival words, are they related to the origin word, when relying on a given semantic network? (In other words, would a semantic word be a good predictor of the arrival of words?). ***[tout un blabla sur les distances dans les substitutions.]***

4.2 Known psycholinguistic effects

We first looked at well-known (and well-studied) psycholinguistic features, and doing so we were able to determine which of the two alternative hypotheses presented in the introduction is valid; the features we examined were age of acquisition norms[7] and number of phonemes[8]. ***[repetition]*** First, by looking at substitution susceptibilities for those features, we can see that words with lower age of acquisition (figure 0**??**) and words with lower number of phonemes (figure 0**??**) are more likely to be substituted than words with higher values for either of those features. This shows us that, although words learned later in development (as well as words with larger number of phonemes) may be cognitively harder to recall, they seem to be substituted according to hypothesis [X]: their relevance and specificness makes them less susceptible to change, whereas words learned earlier or with smaller number of phonemes are more easily substituted.

Now looking at what type of arrival words are selected upon substitution, we can see that both words learned earlier in development (figure 0**??**) as well as words with lower number of phonemes (figure 0**??**) are substituted for words with roughly the same feature values (they tend to augment a little, but stay way below the values expected under $\mathcal{H}_0$ or $\mathcal{H}_{00}$). Conversely, words learned later in development and words with higher number of phonemes, when substituted, go towards words with lower feature values, but no lower than what the null hypotheses predict. This shows again that words learned early on are easily interchangeable for other words learned at about the same age. On the other hand, it seems the cognitive system puts little constraint on the arrival word when substituting a word learned later in development (that is, those words are rarely substituted, but when a substitution does occur the arrival word does not show constraints related to the features we examined). We

[4] To avoid any auto-correlation effect due to the number of substitutions in a cluster (possibly leading to an overly optimistic estimation of confidence intervals), we first average substitutions over each cluster, by considering the average of arrival word features for a given start word. Indeed, substitutions occurring in the same cluster are likely not statistically independent.

[5] Note that $w^\star_{\mathcal{H}_0}$ is in fact constant in this averaging, since by definition it does not depend on w.

[6] In this case $w^\star_{\mathcal{H}_{00}}$ does depend on w.

[7] The Age-of-Acquisition data is obtained from **?**.

[8] The number of phonemes are obtained from the CMU Pronouncing Dictionary for U.S. English [Weide 1998], included in the NLTK package [Bird et al. 2009]. We note that these two measures are likely to be linked, since words with larger numbers of phonemes are likely to be learned later in development.

also tested these results for dependence on the grammatical category of words, i.e. POS tags, and found no effect.

Though not originating from the better-known image-naming or controlled lexical-retrieving tasks, these results shed a new light on the strength of the effect of these features depending on their exact value, and on how these features behave in the context of *substitution* and not only free recall.

4.3 Epidemiological setting

Secondly, and coming back to our primary goal of providing a first empirical testing means for epidemiological models of culture, we added to these two features a number of more abstract properties computed from the Free Association norms used by Griffiths et al. [2007]: namely, the PageRank, betweenness coefficient and clustering coefficient of the words. **[*repetition*]**

These new features are classical measures for network-related matters, but – to our knowledge – have seldom been applied to word characterization. They are only a few among many other features that can be used to characterize words based on the network they form (be it the Free Association norms network, another semantic network, or even a phonological network). We consider the complete set of these features (which includes the two mentioned in the previous section, the new network-based ones, as well as any other network-based feature one can compute) as traits characterizing words in an empirical epidemiological evolution of quotes. The benefit we get from measuring the evolution of such features is in how we can use this information to falsify epidemiological culture models. Indeed, in such a setting the information on how a feature is modified upon substitution (detailed to the individual feature value) is in fact a *fitness landscape* for that particular feature. This fitness landscape can in turn be re-injected into epidemiological models to see if those models account for the empirical distribution of quotes, their life-span and relative success. The data for PageRank, betweenness coefficient and clustering coefficient are shown in figures 0**??**, 0**??** and 0**??**.

[*some talk about what this confirms in Griffith's work*]

Moreover, the substitution susceptibility corresponds to a *mutation probability* in this setting (shown in figures 0**??** for PageRank, 0**??** for betweennes coefficients, and 0**??** for clustering coefficient). To sum up, the data obtained here are empirical measures of (some of) the parameters of epidemiological models of cultural evolution, i.e. essential information to allow empirical testing of these models. If quotes, as a first example of cultural representation, do indeed follow epidemiological rules in their evolution, we may be able to empirically prove the validity – or falsify – the existing models of cultural evolution.

[*this is very rough and brute-like. It needs to be better presented and further developed.*]

[*A big question to be addressed is the "why": what underlying aspect of words with these features make them words to be substituted or not, and why does the brain do that? How was that selected for in evolution?*]

5. PROCESSUAL MODEL

[*Evaluation of the predictability of substitutions for neighbors only, choosing candidates among the immediate neighborhood, proportionally to a score computed over subsets of the above cues (including the empty set as a random baseline), comparing the efficiency of FA and WN in this task.*]

6. CONCLUSION

Main contributions:

—large *in vivo* psycholinguistics experiment, emphasizing the importance of the semantic network structure (Wordnet: Pagerank, degree, clustering[, distance?]), finely describing the impact of classically-influent psycholinguistic variables (*aoa*, number of phonemes, etc.);

—new in the sense that it does not focus on ease of recall but rather bias of substitution, and that in this respect it not only provides a finer description of the bias but also corresponds to an "input-output" reformulation couple describing the joint properties of (substituted→substituting) terms.

—beyond that, provide the first bricks of an empirical *fitness landscape* for the epidemiology of representations

Acknowledgements

We are warmly grateful to Ana Sofia Morais for her precious feedback and advice on this research.

REFERENCES

Scott Atran. Théorie cognitive de la culture. *L'Homme*, 166(2): 107–143, 2003.

Robert Aunger, editor. *Darwinizing Culture: The Status of Memetics as a Science*. Oxford University Press, Oxford, 2000.

S. Bird, E. Klein, and E. Loper. *Natural language processing with Python*. O'Reilly Media, Incorporated, 2009.

Maurice Bloch. A well-disposed social anthropologist's problems with memes. In Robert Aunger, editor, *Darwinizing Culture: The Status of Memetics as a Science*, chapter 10, pages 189–203. Oxford University Press, 2000.

Robert Boyd and Peter J. Richerson. *Culture and the Evolutionary Process*. University of Chicago Press, 1985.

Kit Ying Chan and Michael S Vitevitch. Network structure influences speech production. *Cogn Sci*, 34(4):685–97, 2010. doi: {10.1111/j.1551-6709.2010.01100.x}.

Nicolas Claidière and Dan Sperber. The role of cultural attraction in cultural evolution. *Journal of Cognition and Culture*, 7(1): 89–111, 2007. doi: {10.1163/156853707X171829}.

Jean-Philippe Cointet and Camille Roth. Socio-semantic Dynamics in a Blog Network. In *2009 International Conference on Computational Science and Engineering*, pages 114–121, 2009. doi: {10.1109/CSE.2009.105}.

Allan M Collins and Elizabeth F Loftus. A spreading-activation theory of semantic processing. *Psychological review*, 82(6):407, 1975.

Richard Dawkins. *The Selfish Gene*, chapter 11, pages 189–201. Oxford University Press, 1976. "Memes: The New Replicator".

Paul R Ehrlich and Simon A Levin. The evolution of norms. *PLoS Biol.*, 3(6):e194, 2005. doi: {10.1371/journal.pbio.0030194}.

Linton C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40:35–41, 1977.

Thomas L Griffiths, Mark Steyvers, and Alana Firl. Google and the mind: predicting fluency with PageRank. *Psychol Sci*, 18 (12):1069–76, 2007. doi: {10.1111/j.1467-9280.2007.02027.x}.

Daniel Gruhl, R. Guha, David Liben-Nowell, and Andrew Tomkins. Information Diffusion Through Blogspace. In *Proceedings of the 13th International World Wide Web Conference (WWW'04)*, pages 491–501, 2004.

Adam Kuper. If memes are the answer, what is the question? In Robert Aunger, editor, *Darwinizing Culture: The Status of Memetics as a Science*, pages 180–193. Oxford University Press, 2000.

V. Kuperman, H. Stadthagen-Gonzalez, and M. Brysbaert. Age-of-acquisition ratings for 30 thousand english words. *Behavior Research Methods*, 2012.

Jure Leskovec, Lars Backstrom, and Jon Kleinberg. Meme-tracking and the Dynamics of the News Cycle. *KKD'09*, (June 28-July 1):497–505, 2009a.

Jure Leskovec, Lars Backstrom, and Jon Kleinberg. MemeTracker: tracking news phrase over the web. `http://memetracker.org/`, 2009b. Retrieved on August 19, 2012.

D.L. Nelson, C.L. McEvoy, and T.A. Schreiber. The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods*, 36(3):402–407, 2004.

Lyndsey Nickels and David Howard. Dissociating effects of number of phonemes, number of syllables, and syllabic complexity on word production in aphasia: It's the number of phonemes that counts. *Cognitive Neuropsychology*, 21(1):57–78, 2004. doi: 10.1080/02643290342000122. URL `http://www.tandfonline.com/doi/abs/10.1080/02643290342000122`. PMID: 21038191.

Elisa Omodei, Thierry Poibeau, and Jean-Philippe Cointet. Multi-level modeling of quotation families morphogenesis. In *Proc. ASE/IEEE 4th Intl. Conf. on Social Computing "SocialCom 2012"*, 2012.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report 1999-66, November 1999. URL `{http://ilpubs.stanford.edu:8090/422/}`. Previous number = SIDL-WP-1999-0120.

A. Rey, A.M. Jacobs, F. Schmidt-Weigand, and J.C. Ziegler. A phoneme effect in visual word recognition. *Cognition*, 68(3): B71–B80, 1998.

Beatrice Santorini. *Part-of-Speech Tagging Guidelines for the Penn Treebank Project*, 3rd revision, 2nd printing edition, 1990.

Matthew Simmons, Lada Adamic, and Eytan Adar. Memes Online: Extracted, Subtracted, Injected, and Recollected. In Nicolas Nicolov and James G. Shanahan, editors, *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. The AAAI Press, 2011.

Dan Sperber. *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell Publishers, 1996.

Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

RL Weide. The cmu pronunciation dictionary, release 0.6, 1998.

WordNet. Princeton University "About WordNet.". `http://wordnet.princeton.edu`, 2010. Retrieved on August 19, 2012.

A.P. Yonelinas. The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, 46(3): 441–517, 2002.

J.D. Zevin and M.S. Seidenberg. Age of acquisition effects in word reading and other tasks. *Journal of Memory and language*, 47 (1):1–29, 2002.

APPENDIX

A. FEATURE FIGURES

A.1 aoa Kuperman

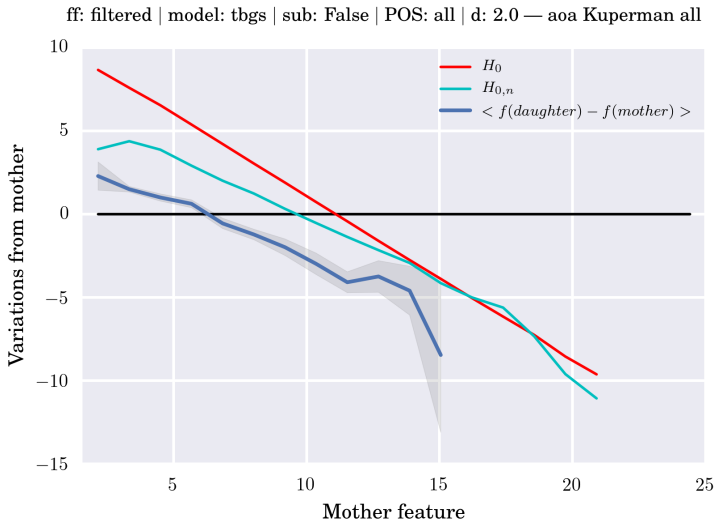
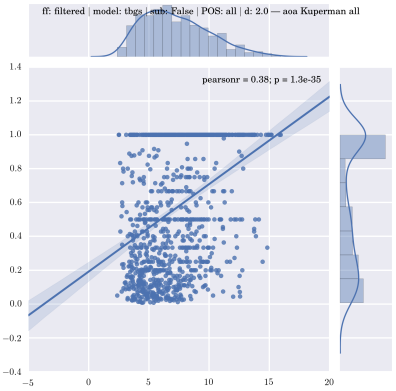
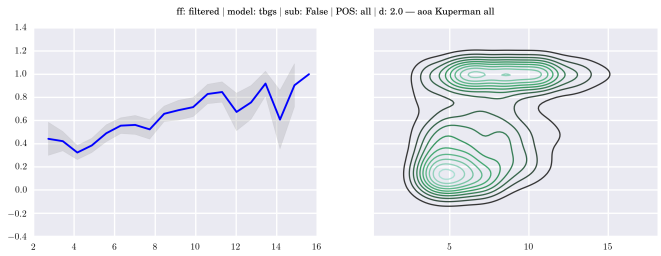


Fig. 2: Feature variation on substitution



(a) Scatter plot



(b) Bin averages and KDE

Fig. 3: Substitution susceptibility

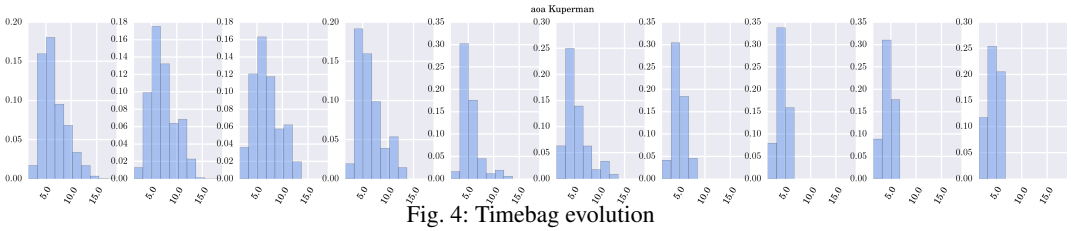


Fig. 4: Timebag evolution

A.2 cmu MNphonemes

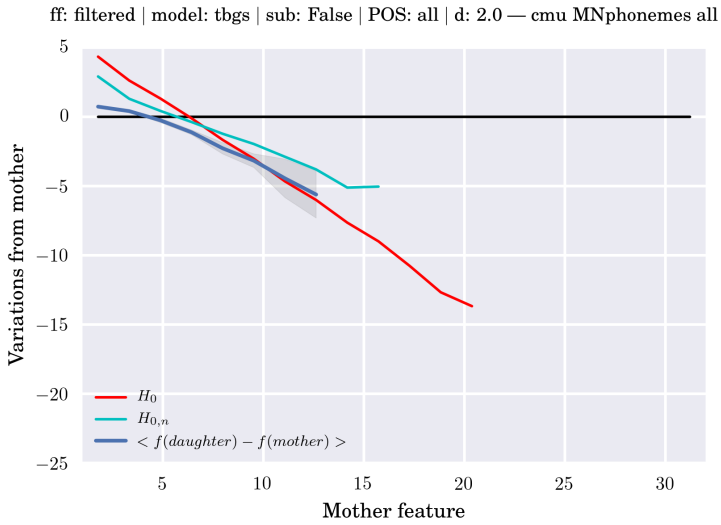
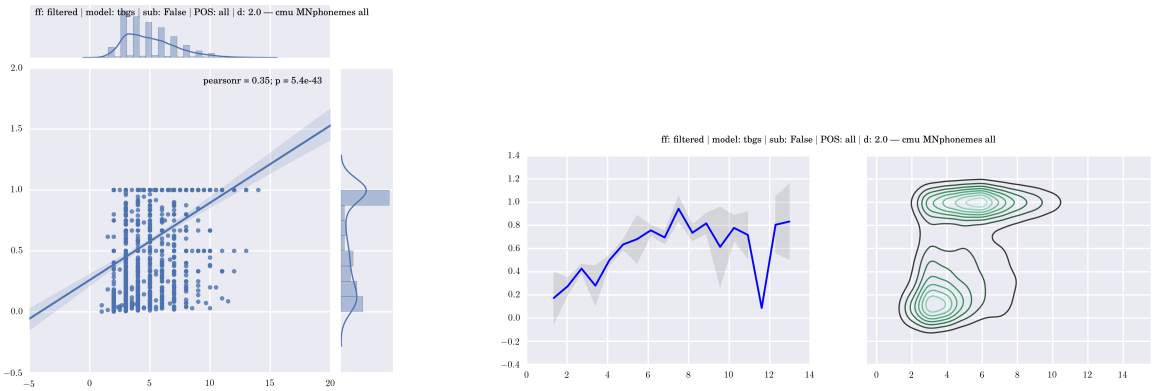


Fig. 5: Feature variation on substitution



(a) Scatter plot

(b) Bin averages and KDE

Fig. 6: Substitution susceptibility

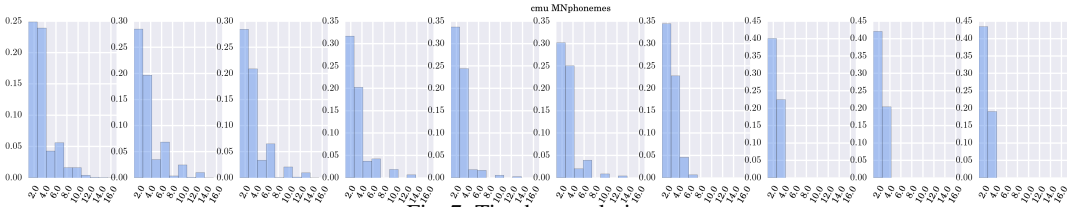


Fig. 7: Timebag evolution

A.3 cmu MNsyllables

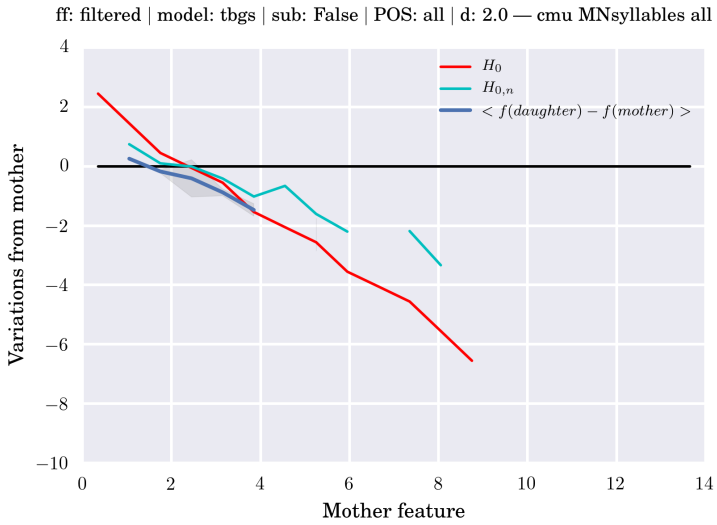
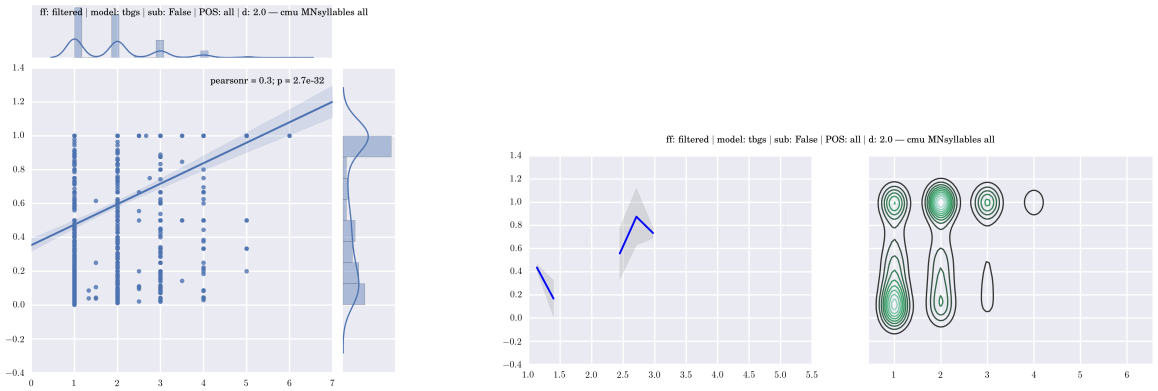


Fig. 8: Feature variation on substitution



(a) Scatter plot

(b) Bin averages and KDE

Fig. 9: Substitution susceptibility

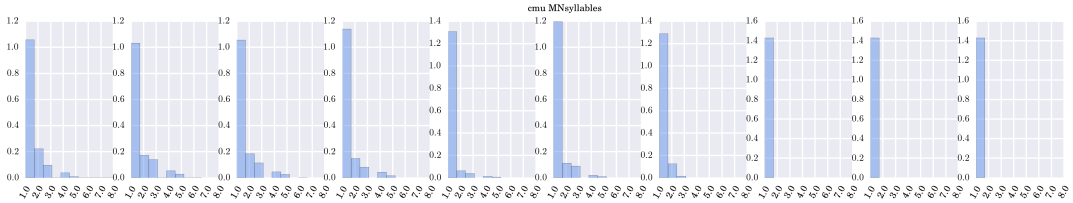


Fig. 10: Timebag evolution

A.4 fa BCs

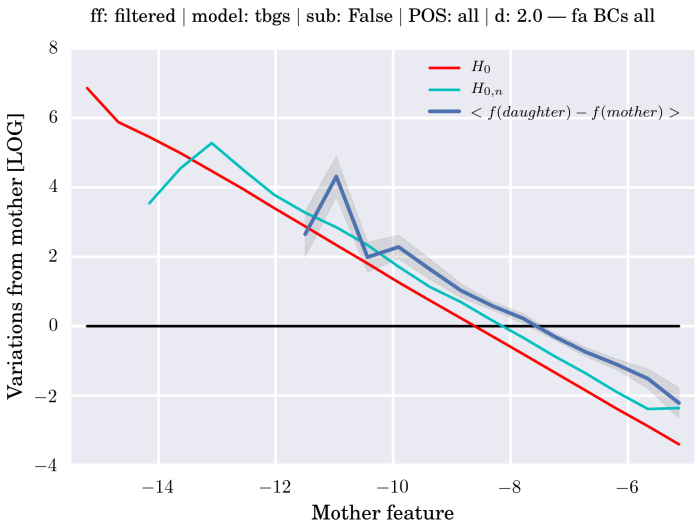
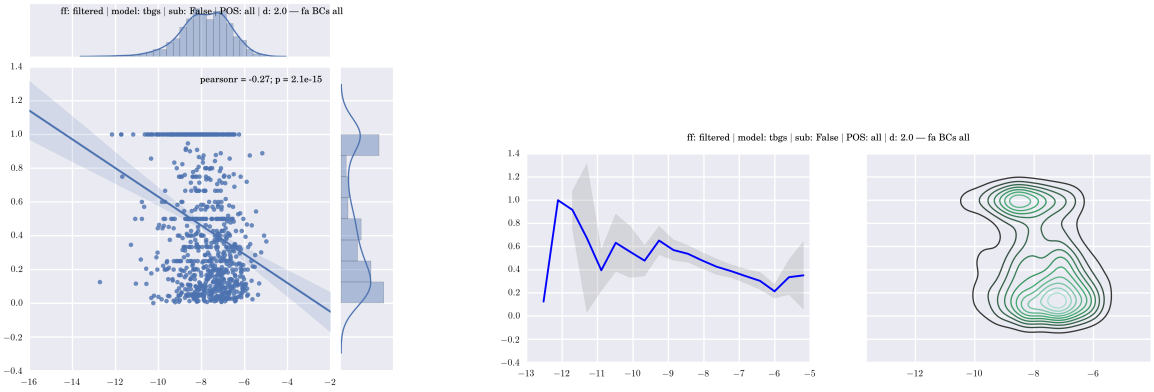


Fig. 11: Feature variation on substitution



(a) Scatter plot

(b) Bin averages and KDE

Fig. 12: Substitution susceptibility

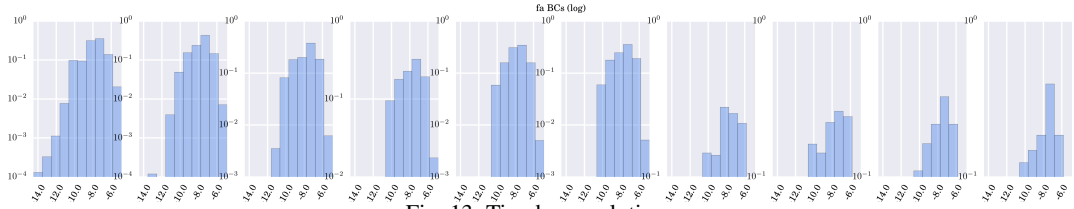


Fig. 13: Timebag evolution

A.5 fa CCs

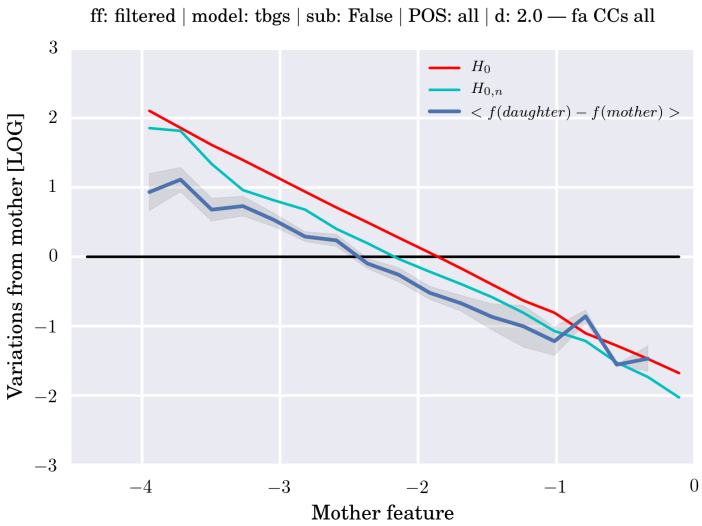
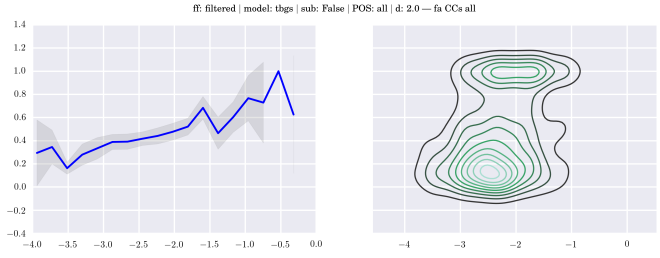


Fig. 14: Feature variation on substitution

(a) Scatter plot



(b) Bin averages and KDE

Fig. 15: Substitution susceptibility

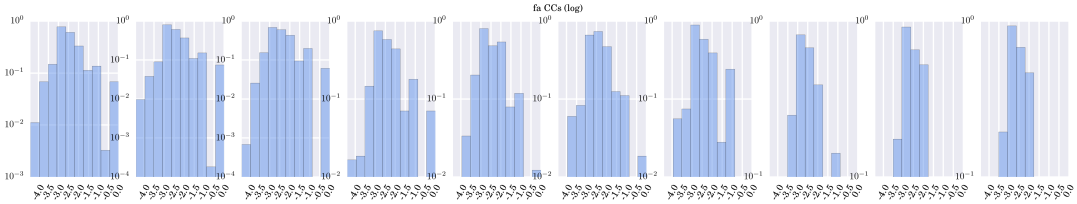


Fig. 16: Timebag evolution

A.6 fa degrees

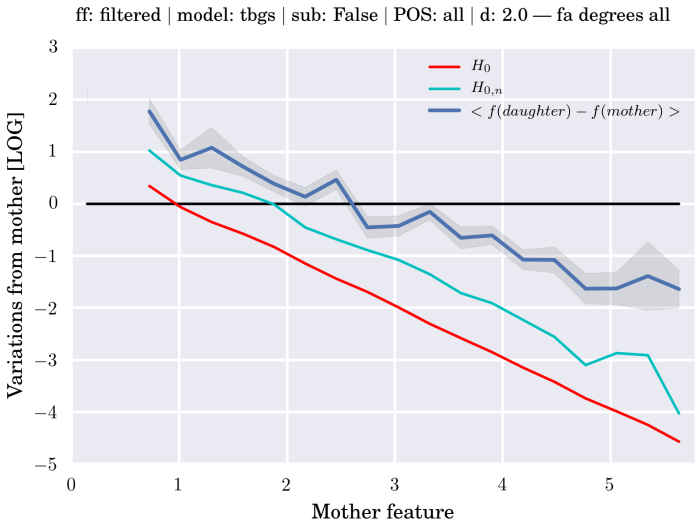
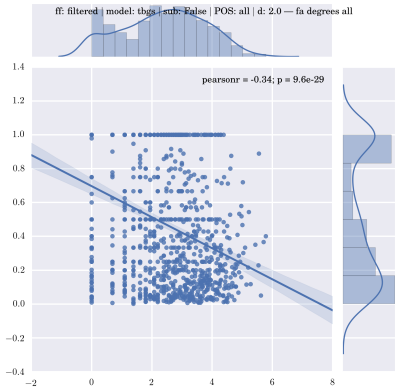
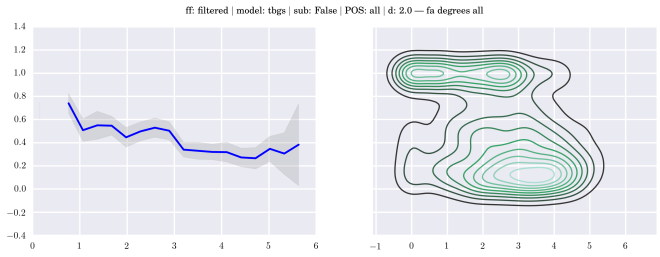


Fig. 17: Feature variation on substitution



(a) Scatter plot



(b) Bin averages and KDE

Fig. 18: Substitution susceptibility

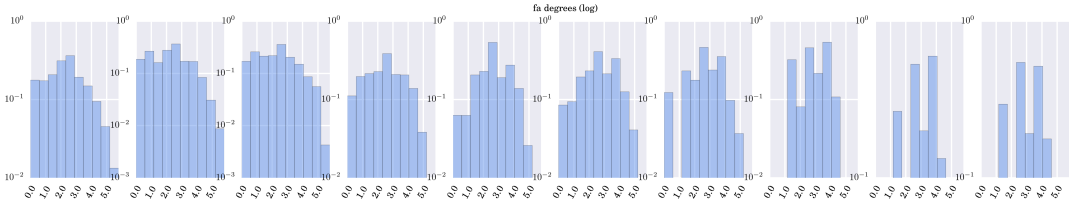


Fig. 19: Timebag evolution

A.7 fa PR scores

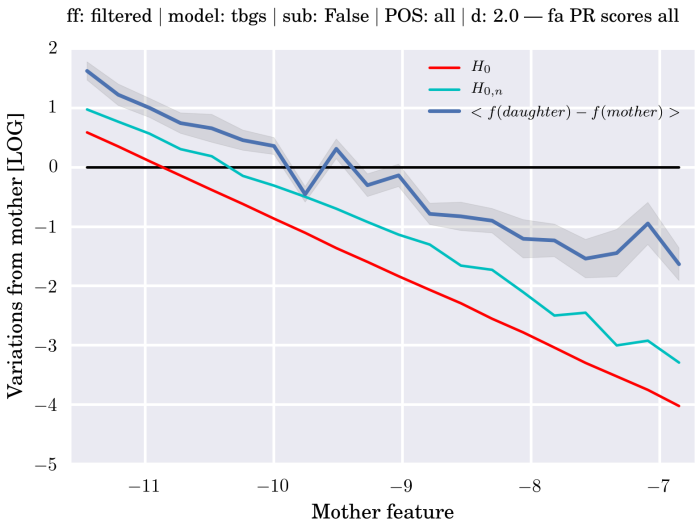
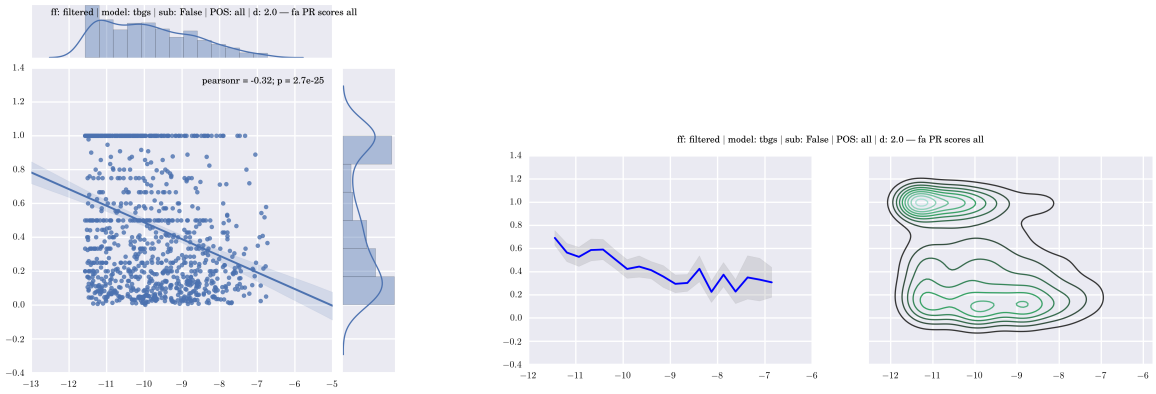


Fig. 20: Feature variation on substitution



(a) Scatter plot

(b) Bin averages and KDE

Fig. 21: Substitution susceptibility

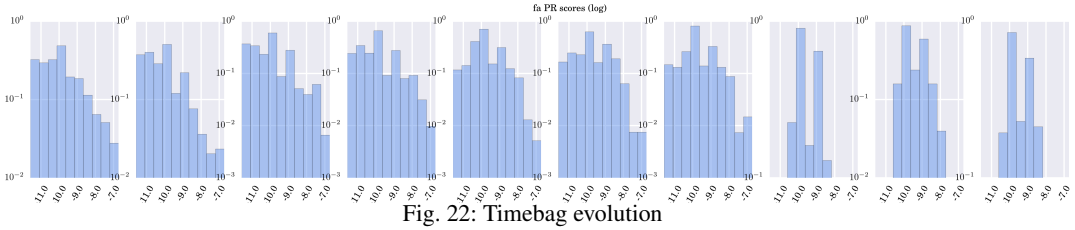


Fig. 22: Timebag evolution

A.8 mt frequencies

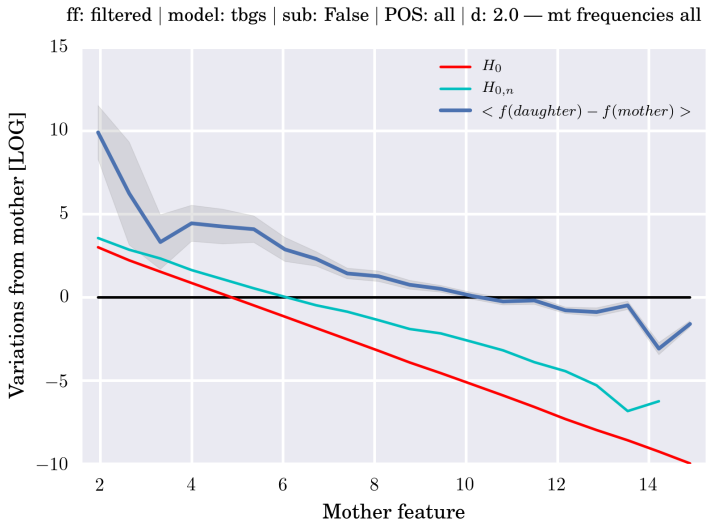
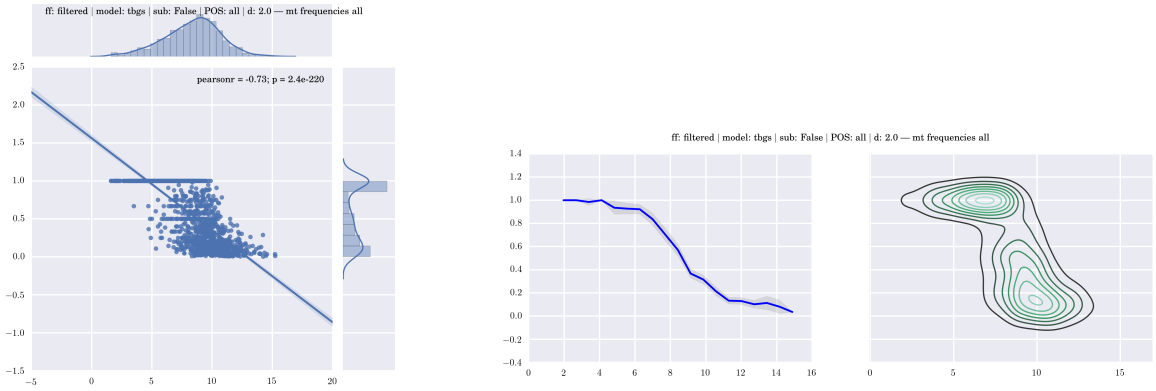


Fig. 23: Feature variation on substitution



(a) Scatter plot

(b) Bin averages and KDE

Fig. 24: Substitution susceptibility

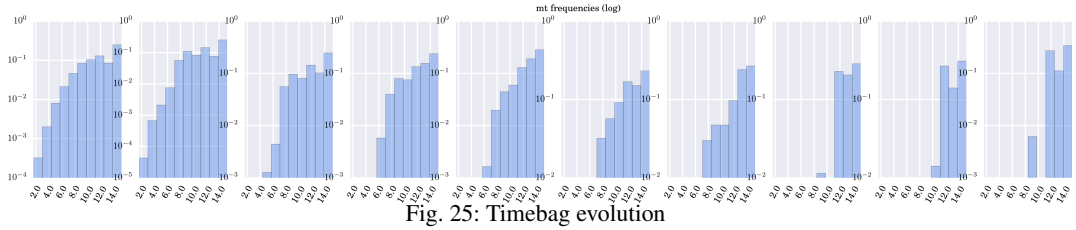


Fig. 25: Timebag evolution

A.9 mt start frequencies

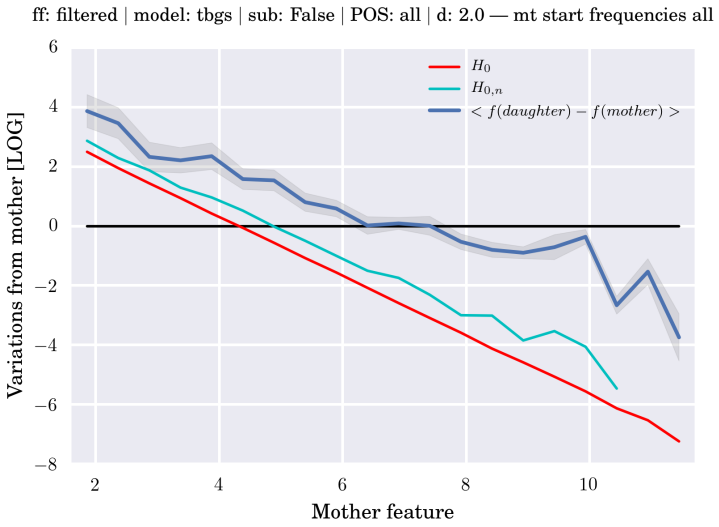
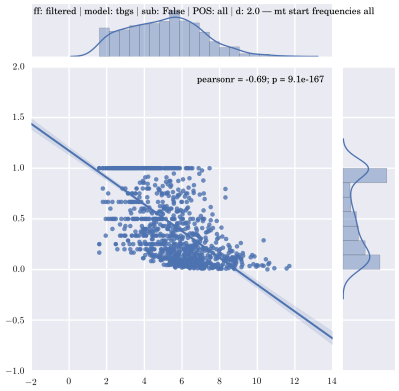
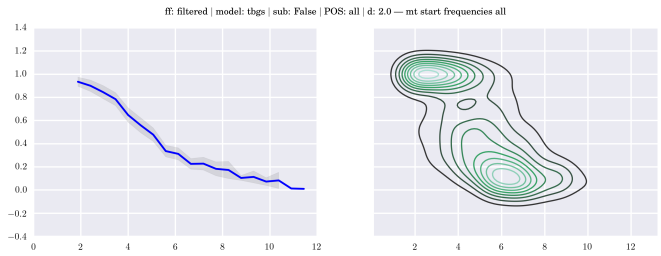


Fig. 26: Feature variation on substitution



(a) Scatter plot



(b) Bin averages and KDE

Fig. 27: Substitution susceptibility

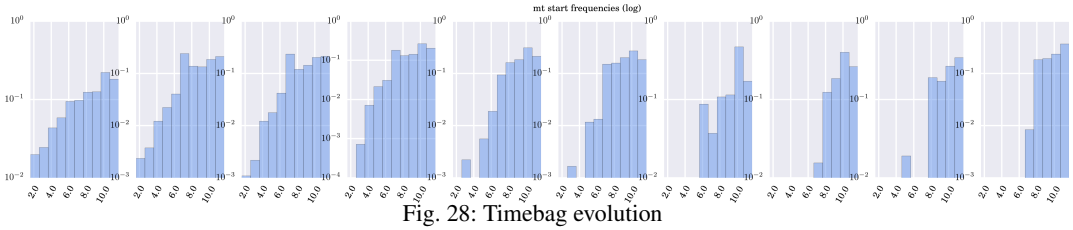


Fig. 28: Timebag evolution

A.10 wn BCs

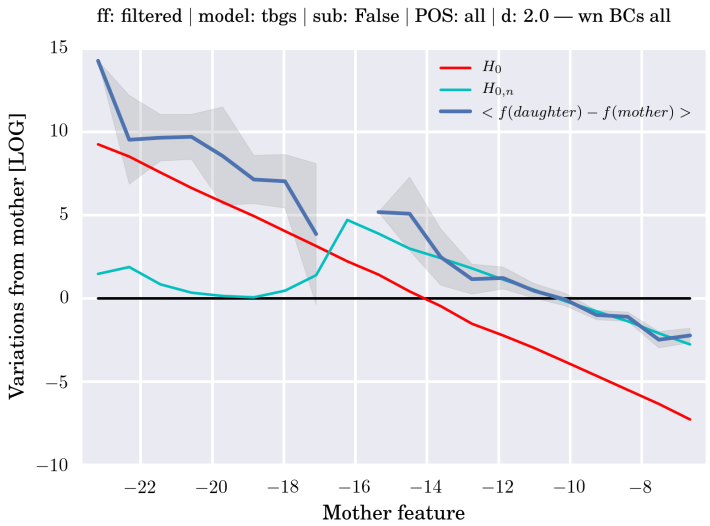
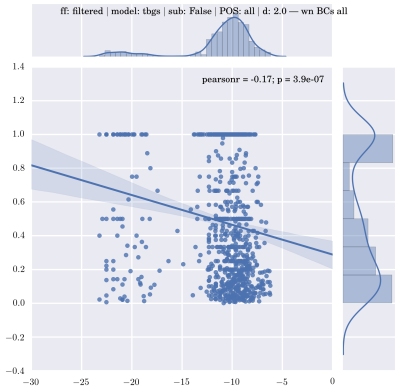
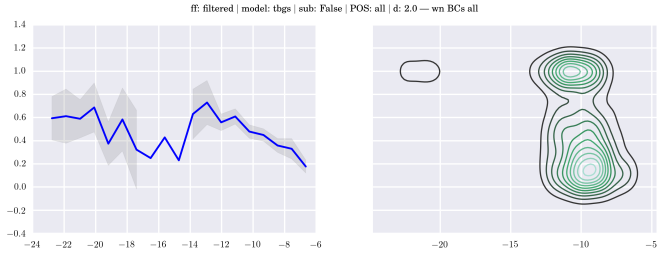


Fig. 29: Feature variation on substitution



(a) Scatter plot



(b) Bin averages and KDE

Fig. 30: Substitution susceptibility

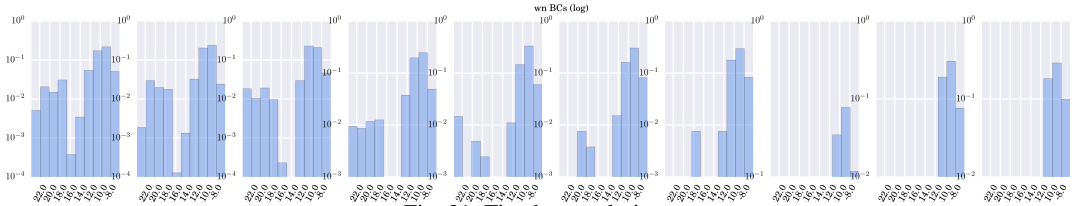


Fig. 31: Timebag evolution

A.11 wn CCs

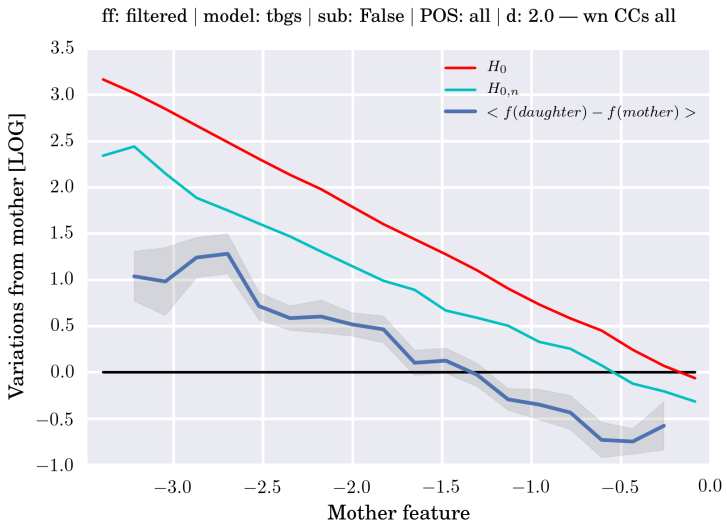
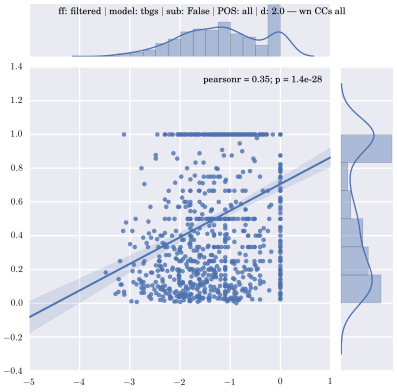
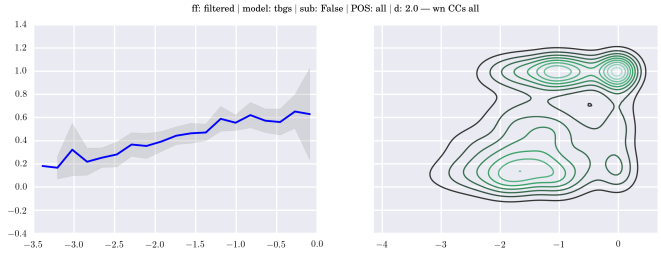


Fig. 32: Feature variation on substitution



(a) Scatter plot



(b) Bin averages and KDE

Fig. 33: Substitution susceptibility

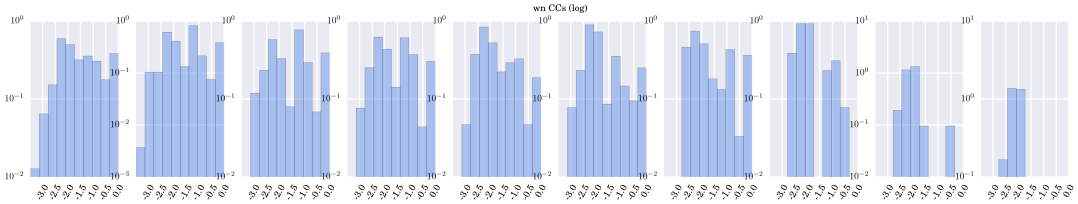
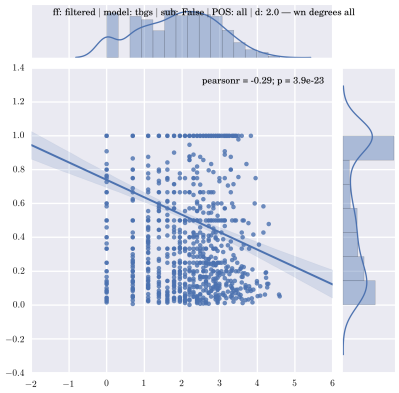


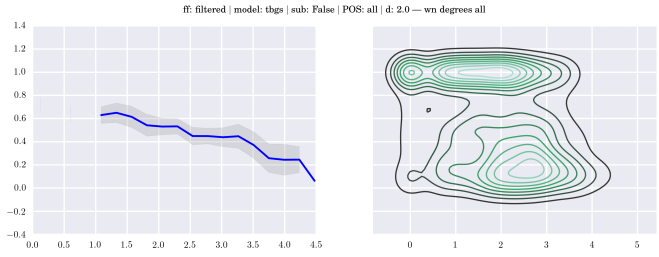
Fig. 34: Timebag evolution

A.12 wn degrees

Fig. 35: Feature variation on substitution



(a) Scatter plot



(b) Bin averages and KDE

Fig. 36: Substitution susceptibility

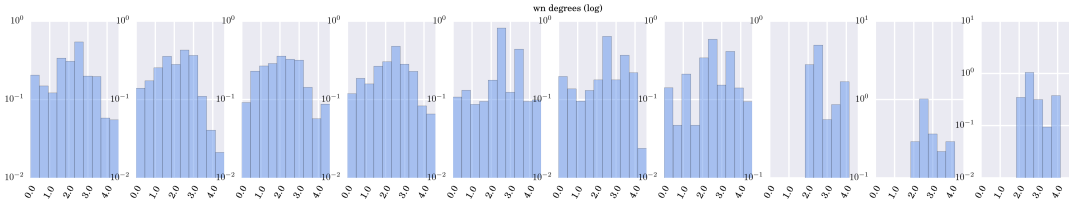


Fig. 37: Timebag evolution

A.13 wn MNSyns

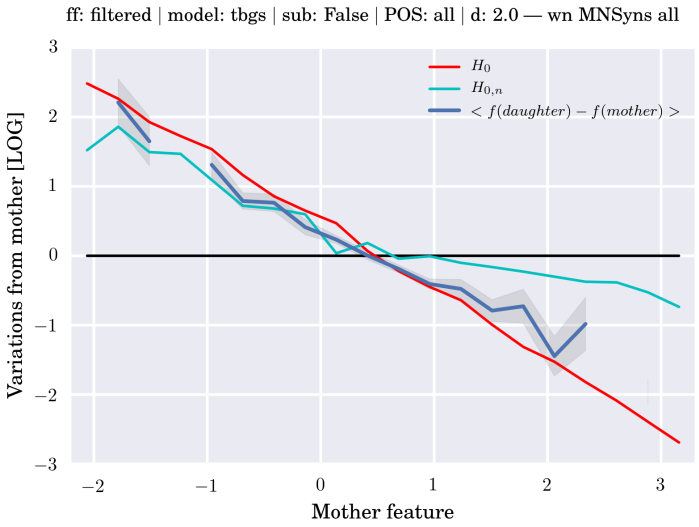
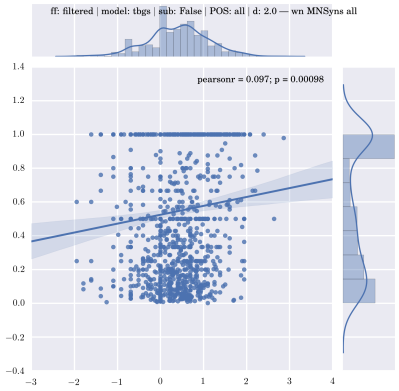
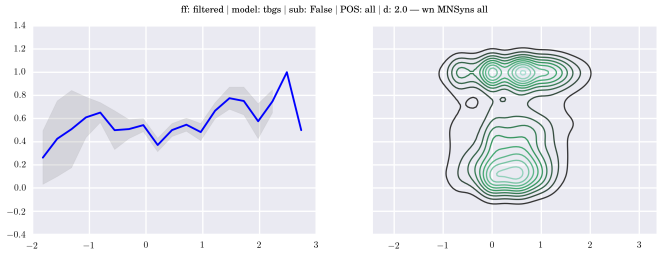


Fig. 38: Feature variation on substitution



(a) Scatter plot



(b) Bin averages and KDE

Fig. 39: Substitution susceptibility

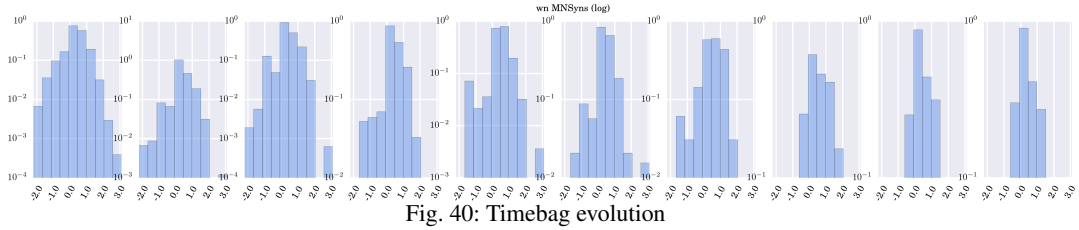


Fig. 40: Timebag evolution

A.14 wn NSigns

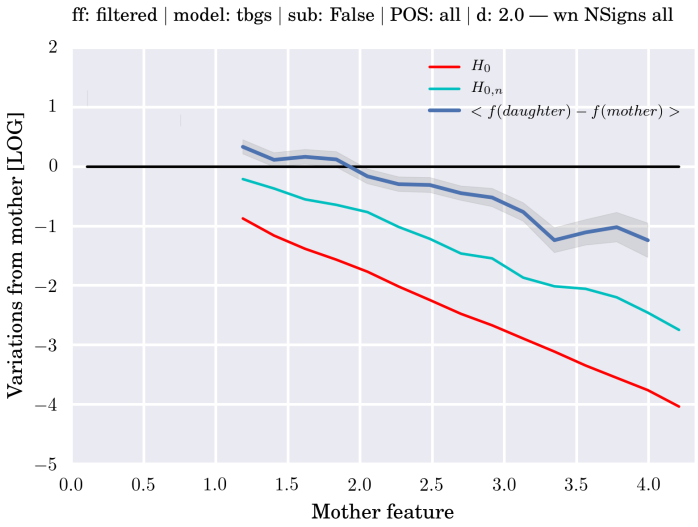
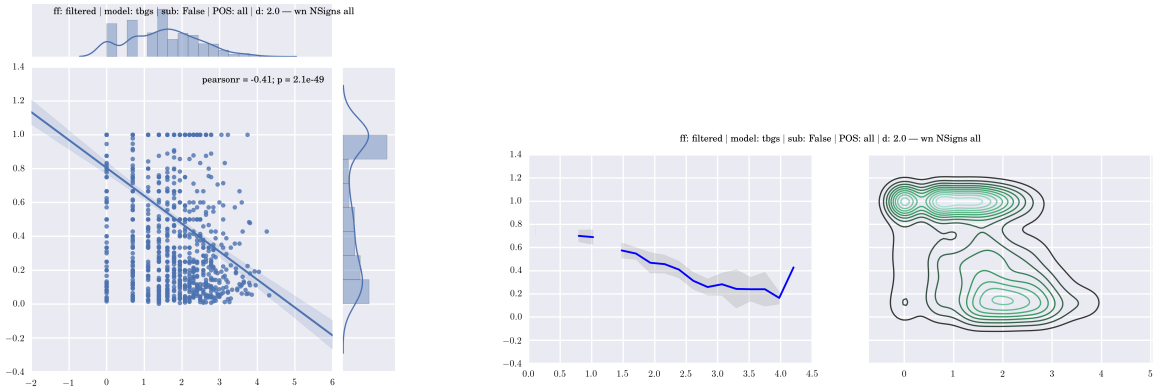


Fig. 41: Feature variation on substitution



(a) Scatter plot

(b) Bin averages and KDE

Fig. 42: Substitution susceptibility

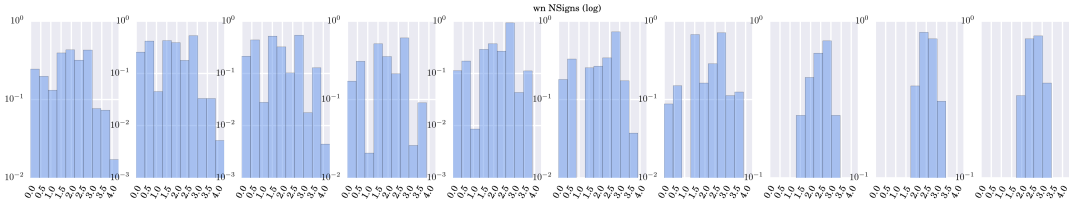


Fig. 43: Timebag evolution

A.15 wn PR scores

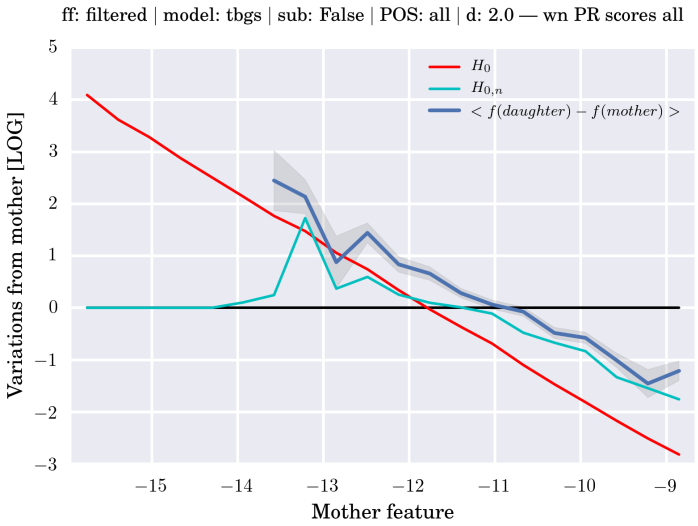
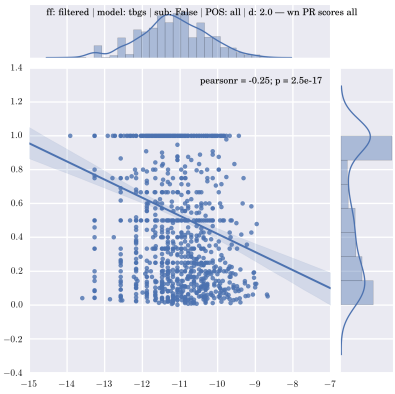
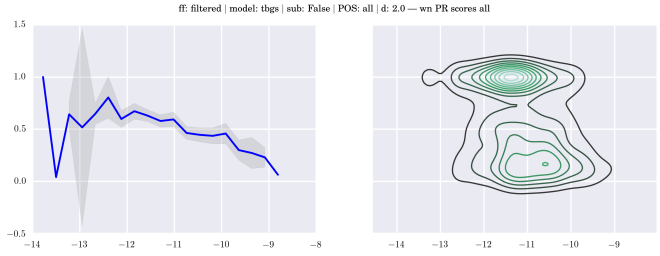


Fig. 44: Feature variation on substitution



(a) Scatter plot



(b) Bin averages and KDE

Fig. 45: Substitution susceptibility

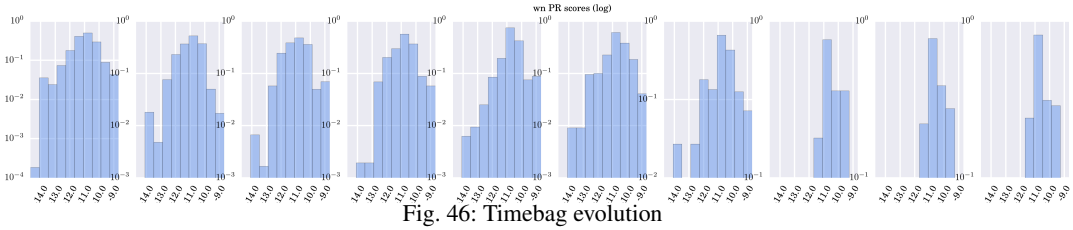


Fig. 46: Timebag evolution