Expert Systems with Applications 40 (2013) 4600–4608

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Complex network tools in building expert systems that perform framing analysis

Erika Fille T. Legara [a,b,*], Christopher P. Monterola [a,b,*], Clarissa David [c]

[a] Complex Systems Programme, Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore
[b] National Institute of Physics, University of the Philippines, Diliman, Quezon City, Philippines
[c] College of Mass Communication, University of the Philippines, Diliman, Quezon City, Philippines

ARTICLE INFO

Keywords:
Framing analysis
Complex network analysis
Clustering analysis
Community detection

ABSTRACT

Framing, in its specific application to media research, is defined as the "central organizing idea for making sense of an issue or conflict and suggesting what is at stake." It can be found in various disciplines of the social sciences, most notably in political science, psychology, and communication research. Due to the fuzzy nature of frames, identifying them has proven to be quite complex. Here, we perform framing analysis on a corpus of news texts on the population and family planning issue in the Philippines by operating two varying approaches: human-based and computer-assisted. A singular holistic approach to framing is initially implemented where coders/domain experts classify each news text to a specific pre-defined frame. This traditional approach is known to raise serious issues on the reliability and validity of the results mainly due to human's intrinsic biases. To address such issues, we propose a novel technique that synergically combines the method of Matthes and Kohring (2008) and complex networks approach. In our model, the codings of texts are cast as a network of content analytic variables (CAVs). Our proposed method tackles the clustering issue that MK raised, which plagues framing scholars in the quantitative identification of news frames in texts. Moreover, the research is significant on a societal level as it also aims to gain perspective for reasons on the lack of progress in discussions about suitable population policies in most developing countries like the Philippines.

1. Introduction

In recent years, much attention has been given to the study of frames in the field of communication research, particularly focusing on how the news media frame issues in text. Framing analysis is salient in understanding the potential of the media to influence public opinion (Entman, 1991; Bardhan, 2001). The subject has garnered a huge appeal across the communication and media disciplines including, but not limited to, policy, media content, and cultural studies (Joachim, 2003; Shah, Watts, Domke, & Fan, 2002; Hoerl, Cloud, & Jarvis, 2009; Yeo, Park, & Arabi, 2007; Ryan, Caragee, & Meinhofer, 2001).

Framing, in this context, is interpreted as the "central organizing idea for making sense of an issue or conflict and suggesting what is at stake (Gamson, 1989; Koch, 1998)." The main challenge is to identify and asses the latent schemes that emerge from a reporting of an issue and recognize them in the most optimized, reliable, and accurate way.

Early framing studies were particularly qualitative in nature using entirely hermeneutic approaches. As such, most of these analyzes were highly dependent on the experts doing the exegesis. Hence, such subjective procedure could potentially cast doubts on the reliability and accuracy of the interpretation (Scheufele & Scheufele, 2010; Van Gorp, 2005); moreover, issues on replicability plague the evaluation measures.

More recent studies, on the other hand, address these issues using more exact quantitative models (Miller, 1997; Miller, Andsager, & Riechert, 1998). Quantitative framing researchers argue that the methods provided are more precise and straightforward and are repeatable – usually done with the aide of "devices" as frame indicators (Koella, 2003). Most of these techniques are computer-assisted, which is the exact antithesis of the more traditional approach. Typically, the methods utilize frequency-statistics of certain keywords and their loci in the body of the texts (Legara, Monterola, David, & Atun, 2010; Murphy, 2001). Although the approach improves reliability, criticisms regarding its validity have been raised since it has been shown that some infrequently occurring words in the text could actually be "central to the meaning of a text"; and, by filtering in only the most frequently occurring words in the examination, a significant amount of nontrivial information can be lost in the process (Scheufele & Scheufele, 2010; Matthes & Kohring, 2008).

Although much has been done in expanding its domain, existing risks vis-à-vis the reliability and validity of framing protocols are still of major concern. Due to this fuzzy nature of frames, naming and quantifying them have proven to be quite complex, which calls

* Corresponding authors. Present address: Complex Systems Programme, Institute of High Performance Computing, A*STAR Computing Science Fusionopolis, 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Singapore. Tel.: +6591167531; fax: +6564631452.

E-mail addresses: legaraeft@ihpc.a-star.edu.sg (Erika Fille T. Legara), monterolac@ihpc.a-star.edu.sg (C.P. Monterola), clarissa.david@gmail.com (C. David).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.01.064

for a multi-disciplinary approach involving both domain and computation experts.

On the other hand, the science of networks has emerged to be a leading approach in probing signatures of complex systems inherent in both nature and society. Networks have been shown to detect cryptic patterns that contain germane information about such systems, e.g. DNA nuecleotides sequence data (Sinatra, Condorelli, & Latora, 2010), dissimilarity between poem and prose (Roxas & Tapang, 2010). In this study, one of the main challenges in applying network theory is in deciphering the hidden patterns through the use of the most appropriate symbolic representations of the compendium of data. It is important to correctly detect the "fundamental units carrying information" in constructing the networks (Sinatra et al., 2010).

Here, we investigate and improve upon existing framing analysis protocols and tackle them in the context of network theory. The research is relevant on two levels: social and methodological. The procedures performed here utilize text from media coverage of a vehemently debated issue in most, if not all, developing countries – the issue on population, reproductive health, and family planning. The population issue has consternated governments of developing countries when concerns are retracted to issues on poverty and women's health. Particularly, we look into the Philippine news coverage. The Philippines is a developing country that is predominantly Catholic (80% of the population), and is ranked fourth (4) worldwide in terms of total population and population density and has been noted by the United States Census Bureau as a fast growing country (Central Intelligence Agency Online Factbook, 2012).

On the methodological significance, detecting media frames has produced a significant number of theories, techniques, and procedures. However, there are very few studies that compares the consistency of the results generated by these various methods because typically only one framing procedure is used to dissect an issue. Here, we utilize and compare three distinct methods in the study of media framing and compare their consistencies, advantages and procedural differences.

2. Data

This study covers a total of 346 news articles related to issues on population, family planning, reproductive health, and contraception. The period covered runs from 1987 to 2007. Data were taken from three of the most widely circulated broadsheets in the Philippines namely, the Philippine Daily Inquirer (PDI), the Philippine Star (PS), and the Manila Bulletin (MB).

3. Singular holistic approach to frames

Perhaps the most traditional method used to find frames is through the perusal of texts by experts of both the issues under study and the field of communication research. A singular holistic approach (SHA) to framing, which is deductive in nature, was carried out. Here, a single frame was considered as a lone reference to a whole article. Coders defined six frames after methodically familiarizing the topic through a careful review of a small sample of text on the issue. The pre-defined frame themes are as follows: the *population and development* frame (F1), the *family planning as conflict between government and church* frame (F2), the *women's and reproductive health* frame (F3), the *population management threatens morals and values* frame (F4), the *population growth and demographic trends* frame (F5), and *others*. Communication research assistants were then instructed to classify each document in the corpus to a single dominant *frame* (chosen from the six pre-defined ones). Table 1 summarizes the number of articles classified in the six frames.

Table 1
Table summarizes the number of articles classified by domain experts in each of the six frames.

Frame	Number of articles
F1	62 (18%)
F2	62 (18%)
F3	104 (30%)
F4	42 (12%)
F5	38 (11%)
others	38

4. Frames as frame elements approach

In Section 4, frames are treated as a collection of elements comprising an issue, which makes the analysis more guided. Essentially, frames are dissected into several parts that are interconnected to each other. This notion was first introduced by Kohring and Matthes (2002) and the conceptual basis behind the technique stemmed from the widely-accepted definition of framing given by Entman: "the process of highlighting some select aspects of a news or an article to publicize a specific *problem definition*, *causal interpretation*, *moral evaluation*, and/or *treatment recommendation* for the item described" (Entman, 1993). These elements are thought to constitute a frame theme. It should be noted that depending on the issue at hand, there can be multiple cases of each of these elements.

Content analytic variables (CAVs) were then infused in the analysis making the classification more systematic. As an example, for the frame element *problem definition*, the corresponding CAVs are: topic/theme, actor, and proponent. Table 2 summarizes the frame elements and the associated CAVs. These CAVs are further anatomized into single binary variables, resulting to a total of 71 variables (see Tables 3–6). Coders then analyze each article based on the binary CAVs. A sample coding sheet is shown in Table 7 where columns represent the CAV scores while rows represent the cases.

4.1. Network construction of content analytic variables

Since by definition a frame is a collection of frame elements, it is presumed that such consists of a unique set of binary variables that form a dominant pattern, which emerges out of the CAVs' interconnectivity. Matthes and Kohring (2008) emphasized that this procedure has an added complexity since it calls for an optimized clustering procedure of frame elements. In fact, MK disclosed that one of the liabilities in their method of using k-means clustering is that "problems may occur when conducting cluster analysis of frame element" since it does not have a systematic way of determining the optimal number of clusters (Matthes & Kohring, 2008).

One of the strengths of using complex networks in clustering analysis is its ability to extract these latent patterns without prior

Table 2
Frame elements and the corresponding content analytic variables.

Frame elements	Content analytic variables
Problem definition	Topic or theme
	Actor
	Proponent
Causal interpretation	Benefit attribution
	Risk attribution
Moral evaluation	Benefits
	Risks
Treatment recommendation	Solution
	Treatment

Table 3
Content analytic variables defining the element *Problem Definition*.

Frame element	Variable	Description
Topic theme	ttabort	Topic/Theme: Abortion
	ttpublic	Topic/Theme: Public opinion
	ttchurch	Topic/Theme: Church statements on positions
	ttlegis	Topic/Theme: Legislation for reproductive health/population
	ttnfp	Topic/Theme: Natural family planning/Responsible parenthood
	ttgrowth	Topic/Theme: Population growth or trends
	tthunger	Topic/Theme: Food or rice shortage/Hunger
	ttrepro	Topic/Theme: Reproductive health/Maternal & child health
	ttgender	Topic/Theme: Gender or women's rights
	ttteen	Topic/Theme: Adolescent sexuality/Teenage pregnancy/Sex education
	ttaids	Topic/Theme: HIV/AIDS and other sexually transmitted diseases
	ttmethod	Topic/Theme: Family planning methods
	ttgov	Topic/Theme: Government's popn management program
Actor	acchurch	Actor: Church or other religious groups
	acpolicy	Actor: Researchers and policymakers
	aclgu	Actor: Local government official
	acpres	Actor: President/Presidential spokesperson
	acdoh	Actor: Secretary or officials of Department of Health and Population Commission
	acdeped	Actor: Secretary or officials of Department of Education, Department of Agriculture, and other agencies
	acdoctor	Actor: Frontline health service providers
	acintorg	Actor: International organization
	aclegis	Actor: Legislature (House of Representatives and Senate) and its members
	acwngos	Actor: Women/gender interest groups and other non- governmental organizations
Proponent	prchurch	Proponent: Church groups
	prpres	Proponent: Office of the President or the government
	prdoh	Proponent: Department of Health or Popn Commission
	prgovt	Proponent: Other government agencies
	prlgu	Proponent: Local government units
	prlegis	Proponent: Legislature (Congress or Senate)
	prngo	Proponent: NGOs and other interest groups
	prdoctor	Proponent: Medical/health professionals or groups
	prresrch	Proponent: Researchers or research institutions
	printorg	Proponent: International organizations or institutions

Table 4
Content analytic variables defining the element *Moral Evaluation*.

Frame element	Variable	Description
Benefit	bqlife	Benefit: Improved quality of life
	bmhlt	Benefit: Improved maternal and child health care
	bteen	Benefit: Decreased teenage/early pregnancy
	babort	Benefit: Decreased unintended pregnancies and/or abortion
	bdev	Benefit: Economic development
	beduc	Benefit: Improve children's chance of good education
	bdpov	Benefit: Reduce poverty
Risk	rmchlth	Risk to maternal and child health
	rabort	Risk: Contraception is abortive/will increase abortion
	rsin	Risk: Moral and spiritual well-being, committing a "sin"
	rhlth	Risk: Contraception is dangerous to people's health
	recon	Risk: Economic risks (rpov, rdev, runemp)
	resrce	Risk: Depletion of resources, hunger, congestion
	rsocser	Risk: Poor delivery of social services from government (i.e., education, health)
	rsexpreg	Risk: Promote promiscuity, premarital sex, early pregnancy
	rfamily	Risk: Risk to marriage and family life

Table 5
Content analytic variables defining the element *Causal Interpretation*.

Frame element	Variable	Description
C. Causal interpretation		
Benefit attribution	bafp	Benefit attribution: Programs that promote FP/population management
	banatfp	Benefit attribution: Natural family planning/Responsible parenthood only
	baecon	Benefit attribution: Economic development caused population growth to decrease
	bagovt	Benefit attribution: Other government programs
	bacurb	Benefit attribution: Curbing population growth
Risk attribution	rachurch	Risk attribution: Church as hindering efforts to control population growth
	ragovt	Risk attribution: Government's failure or lack of population management programs
	rafplan	Risk attribution: FP programs and contraception promotion
	rapgrow	Risk attribution: Rapid population growth

knowledge about the number of clusters that need to be extracted. This fundamentally addresses the clustering issue raised by MK.

Hence, we construct a network from the corpus $\mathcal{A}$ of $n = 346$ articles, i.e. $\mathcal{A} \equiv \{\alpha_1, \alpha_2, \ldots, \alpha_{n=346}\}$, and then perform contemporary clustering methods from network theory.

Each article α_i in the ensemble is described by a set $\mathcal{V}_i \equiv \{v_1, v_2, \ldots, v_{j=71}\}$ of variable ratings that provides coder ratings for each CAV v_j in $\mathcal{V}_i$. Now, for every article $\alpha_i \in \mathcal{A}$, a 71×71 adjacency matrix $\mathbf{M}_i$ of all content analytic variables $v_j \in \mathcal{V}_i$ is constructed that represents the connectivity of vertices of the weighted graph $\mathcal{G}_i(V_i, E_i)$, where $E_i = \{e_{kl}\}$ is the set of all edges or links in the network and V_i is the set of all nodes representing all 71 analytic variables $\mathcal{V}_i \equiv \{v_1, v_2, \ldots, v_{j=65}\}$ in the study. $\mathbf{M}_i$ is a symmetric *(0,1)*-matrix with diagonal entries $m_{kk} = 0$ since $\mathcal{G}_i$ is an unweighted undirected graph with no self-loops. In $\mathbf{M}_i$, $m_{kl} = 1.0$ if content analytic variables v_k and v_l are rated 1.0 together in article α_i, $m_{kl} = 0$ otherwise. To illustrate, we utilize the schematic coding in Table 7 to construct three networks, one for each article α_i (see Fig. 1). For every $\mathcal{G}_i$, a maximum clique $C_{i,\max}$, which is a subset of V_i exists such that for every two vertices v_k and $v_l \in C_{i,\max}$ an edge $e_{kl} \in E_i$ exists. In Fig. 1, Article 1 has maximum clique $C_1 = \{v_1, v_3, v_4, v_7\}$. The clique vertices show all content analytic variables rated 1.0 together in sample article α_1.

Table 6
Content analytic variables defining the element *Treatment Recommendation*.

Frame element	Variable	Description
Solution	sfplan	Solution: Promote FP and contraception
	snatfp	Solution: Promote natural FP/responsible parenthood
	spopmgt	Solution: Stronger popn management program by govt
	secon	Solution: Economic development
Treatment	positive	Treatment: Positive
	negative	Treatment: Negative
	neutral	Treatment: Neutral

Table 7
Schematic coding of articles α_i where $v_1, v_2, \ldots, v_7$ represent sample content analytic variables. The values are rated 1 or 0.

	v_1	v_2	v_3	v_4	v_5	v_6	v_7
article α_1	1	0	1	1	0	0	1
article α_2	0	0	0	1	0	0	1
article α_3	1	1	1	0	1	1	0

Fig. 2 shows an empirical maximum clique $C_{0,\max} \subset \mathcal{G}_0$ extracted from one of the sample documents in the corpus. In the figure, isolated vertices are removed for a better illustration, thereby leaving a fully-connected network. Constructing all maximum cliques across 346 articles, it was found that the average clique number $\bar{\omega}(\mathcal{G}_i)$ is 11.

Finally, a weighted graph $G(V, E)$ was constructed from the individual $\mathcal{G}_i$'s where, this time, $E(G) = \{e_{kl}\}$ is the set of all *weighted* edges $w_{kl} \in \{\mathbb{Z}^+\}$. In terms of the adjacency matrix representation as discussed previously,

$$\mathbf{M}_G = \sum_{i=1}^{346} \mathrm{M}_i. \tag{1}$$

The final weighted network $G(V, E)$ is shown in Fig. 3.

4.2. Patterns through community detection

The method of community detection was then applied on graph $G(V, E)$ to reveal grouping patterns among CAVs. A community, in the context of complex network theory, is the "existence of the networks' structural subunits that are associated with more highly interconnected parts." Palla, Dernyi, Farkas, and Vicsek (2005) Essentially, natural partitions are identified and extracted within the network through non-homogeneous connectivities. A modularity optimization algorithm introduced by Blondel et al. was utilized (1) because of its capability to return high-quality partitions, which are quantified by the *modularity function* of the communities; and (2) for fast community detection, which was shown to be superior in terms of computation time when compared to other existing

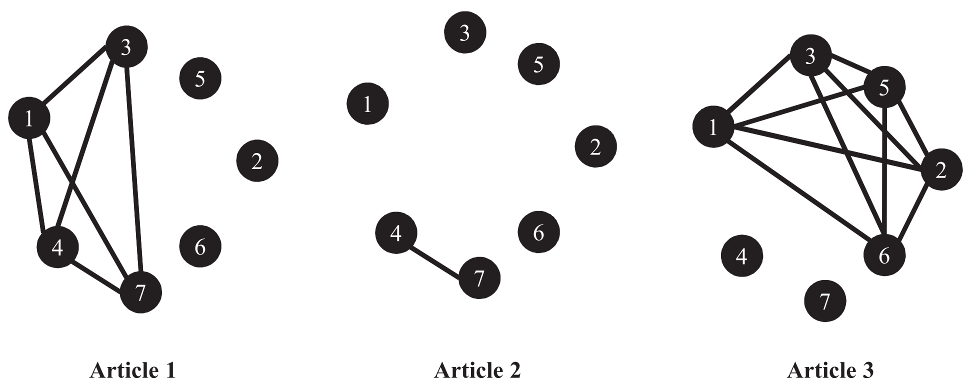

Fig. 1. Three networks $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ constructed from the schematic sample coding in Table 7. Nodes represent the seven (7) content analytic variables in the example.

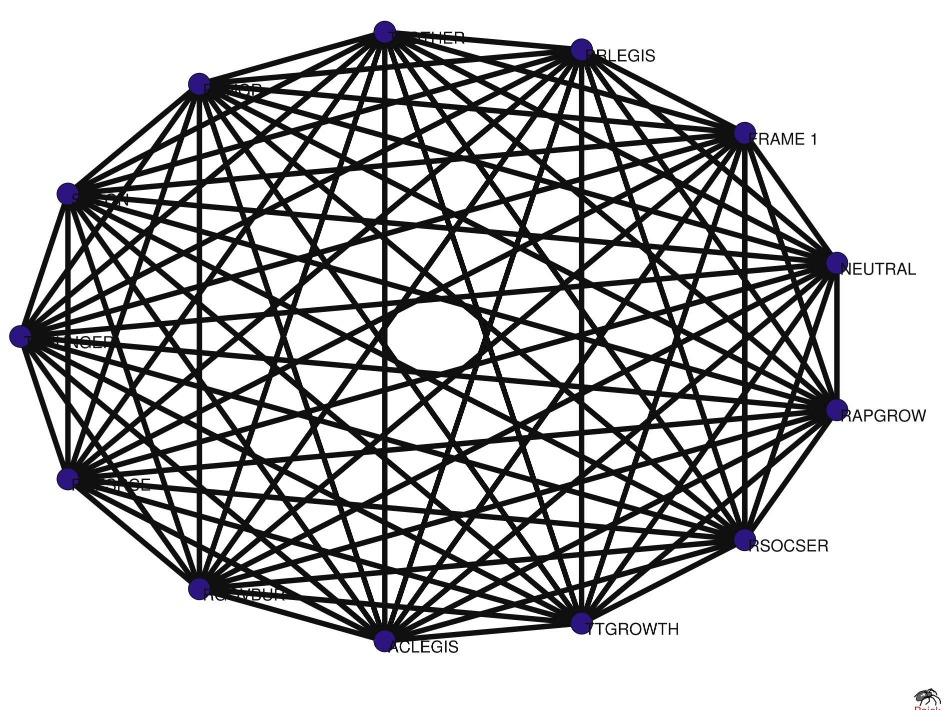

Fig. 2. An example of an empirical maximum clique. Shown are the content analytic variables (CAVs) and their fully connected links, i.e. the CAVs were all rated 1.0 together in a case.

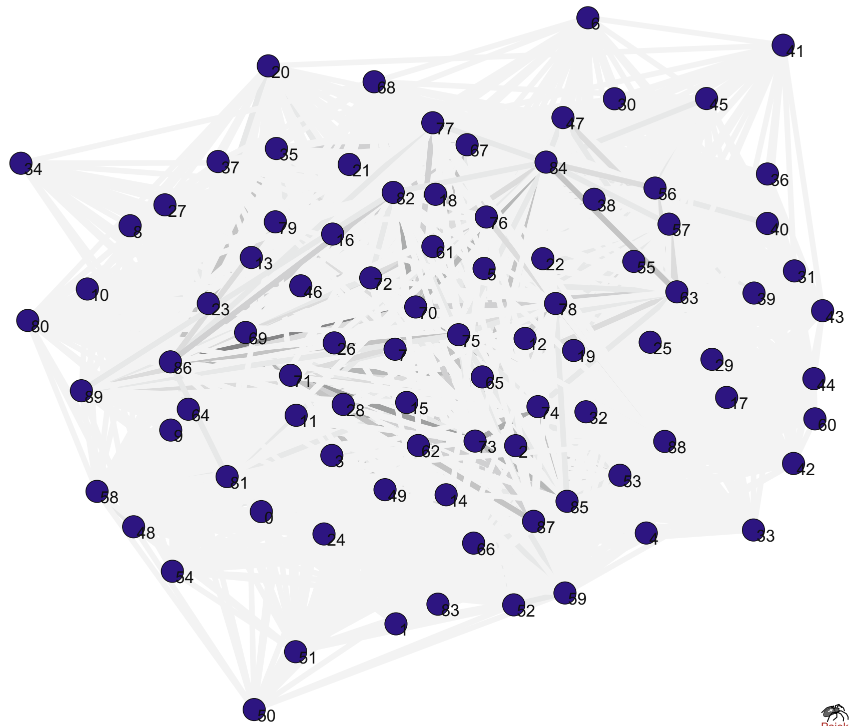

Fig. 3. Weighted network $G(V, E)$ of binary analytic variables from the entire corpus. $E(G) = \{e_{kl}\}$ is the set of all edges weighted $w_{kl} \in \{\mathbb{Z}^+\}$ and $V(G)$ is the set of nodes (or vertices) representing all 65 content analytic variables. The weight of the edges indicate the number of times two variables k and l were rated 1.0 together in the corpus.

algorithms (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008). The modularity Q of a partition evaluates the density of edges within communities as compared to edges between communities (Newman, 2006; Newman & Girvan, 2004). The value lies in the range $[-1,1]$. Mathematically, it is defined as

$$Q = \frac{1}{2m}\sum_{k,l}\left[w_{kl} - \frac{s_k s_l}{2m}\right]\delta(c_k, c_l), \tag{2}$$

for weighted networks where w_{kl} represents the edge weight between vertices k and l, $s_k = \sum_l w_{kl}$ is the sum of all the edge weights connected to vertex k, c_k is the community to which vertex k is assigned, the $\delta(c_k = c_l) = 1$ if $c_k = c_l$ and 0 otherwise and $W = \frac{1}{2}\sum_{kl} w_{kl}$ (Newman & Girvan, 2004).

The greedy-algorithm introduced by Blondel et al. is iterative and involves two general steps. First, it goes through all vertices in sequence. It starts, for example, at vertex k. It then calculates for the weighted modularity Q using Eq. (2) every time vertex k is placed in the community of its neighbor l. It goes through all the neighbors l and calculates for Q. It then picks out the community of l that returns the highest increase in Q, if and only if it is positive. This process results to a first level partition. In the second step, the extracted communities are substituted with dummy vertices called "supervertices". This virtually reduces the network size. Two "supervertices" are then linked together if there exists at least an edge between vertices of the respective communities at the lower level. The edges between these "supervertices" are weighted by adding the weights of the represented partitions at the lower level. The algorithm is described in detail in reference (Blondel et al., 2008).

Finally, we have chosen this particular method over other more traditional clustering algorithms because it automatically returns

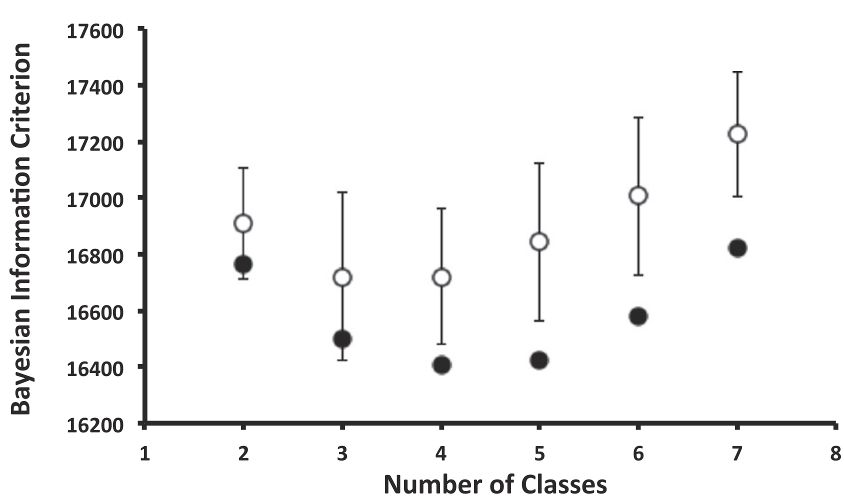

Fig. 4. The figure shows the obtained Bayesian Information Criterion minimum and mean values for k clusters using Latent Class Analysis for cluster analysis.

the optimal number of communities that results to maximum modularity, i.e. the number of partitions are not pre-defined unlike in other clustering methods where one has to indicate the number of partitions beforehand (e.g. *k*-clustering). In terms of maximizing the reliability of framing analysis, analysts need not impose the number of clusters that is expected from the collection; therefore, it is more objective in nature. Another advantage of this algorithm is that there is no unnecessary removal of information caused by the deletion of weak links in the analysis as opposed to other methods where cutoffs have to be set to gain more defined results (Matthes & Kohring, 2008).

4.3. Community detection vs latent class analysis

To demonstrate the superiority of the method over other clustering methods, we perform latent class analysis (LCA), which is regarded by framing scholars as the "most advanced clustering technique" (Personal communication) that gives a clear criteria for the number of clusters of frame elements in the analysis. LCA is especially superior in explorative clustering analysis where no information about the theoretical number of clusters is indicated. The basic latent class model is a finite mixture model (Lazarsfeld, 1950; Lewis & Linzer, 2011) that links a series of manifest variables to a group of latent classes. In this research, LCA was implemented in **R** using the polytomous variable latent class analysis (poLCA) **R** package developed by Lewis and Linzer (2011) where expectation–maximization and Newton–Raphson algorithms were utilized to find maximum-likelihood estimates of the manifest variables.

In LCA, the two most commonly used measures that determine the best-fitting model in a system are the Bayesian information criterion (BIC) (Schwartz, 1978) and the Aikaike information criterion (AIC) (Akaike, 1973). The best-fitting model is the one that minimizes the criteria. Without loss of generality, we used the BIC (Forster, 2000; Lin & Dayton, 1997), which is given by

$$\mathtt{BIC} = -2\Gamma + \Phi \ln N, \tag{3}$$

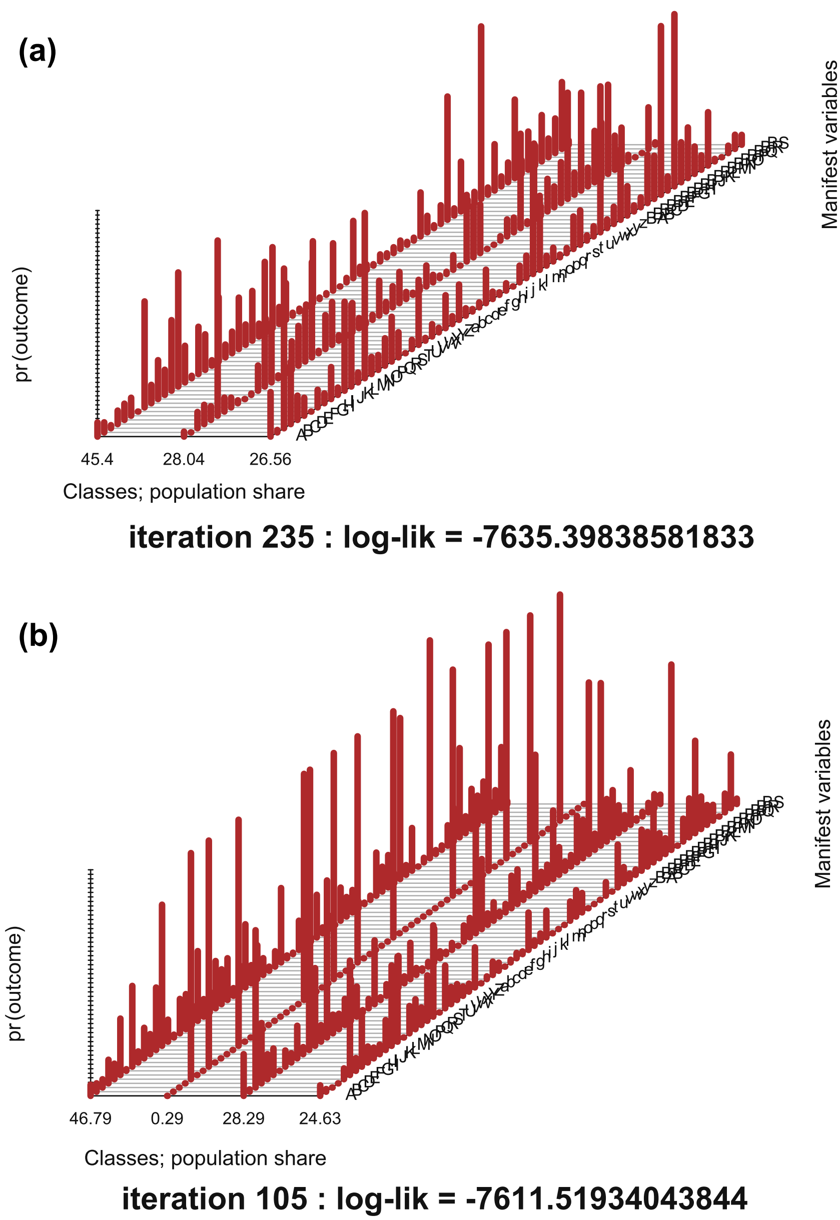

Fig. 5. Sample plots of the outcome probability of the manifest variables and the population shares for the (a) three-class and (b) four-class models in one of the 500 runs implemented.

Table 8
Three communities/partitions of CAVs obtained using community detection algorithm (CDA).

Community 1
runemp, bacurb, rgvhlth, rapgrow, rpov, rhous, acdeped, acpres, secon, rhngr, rsocser, bdev, rgovbur, prgovt, neutral, bdpov, reduc, tthunger, rdev, recon, ttgrowth, rachurch, beduc, bqlife, renvt, rresrce, ttother
Community 2
acdoh, ttteen, ttrepro, prresrch, bafp, ttaids, prngo, prlg, baecon, bagovt, sfplan, acdoctor, ttgender, acpolicy, positive, prdoh, raother, acintorg, aclg, printorg, ttgov, ttpublic, rmchlth, ragovt, bteen, acwngos, bmhlt, prdoctor, spopmgt, babort
Community 3
rfamily, prpres, acchurch, snatfp, sother, rabort, ttchurch, rpmsex, prchurch, prother, baother, rmarr, rtprg, acothers, ttmethod, banatfp, rsexpreg, rsin, prlegis, ttabort, negative, rhlth, ttlegis, ttnfp, rfam, aclegis, rafplan

where Γ is the maximum log-likelihood of the model and Φ the number of evaluated variables.

We did several runs (500) for various classes (k = 2 to 7) in the LCA and calculated for the respective mean and minimum BIC values. Fig. 4 plots the different BIC values obtained, highlighting both the mean (hollow) and minimum (filled) BICs for each class k. The plot indicates that if the minimum BIC is utilized as an indicator, the best model gives four clusters (4 frames). However, if the mean BIC is used, the best model results to three culsters (3 frames). Looking closer into the three (3) and four (4) classification models, respectively, we have Fig. 5(a) and (b). The two graphs indicate the population shares of the classes and the outcome probabilities (strength) of the manifest variables. In the four-class system (Fig. 5(b)), it can be noted that one of the classes has a population share of less than 0.50% of the entire corpus of articles. This partitioning could just as well be a three-class system (Fig. 5(a)) since a share of less than 0.50% is less than the standard deviation of the population distributions across runs.

5. Results and discussion

The results using the community detection algorithm (CDA) indicate that the most optimal partition would be to group the articles into three (3) communities (*see* Table 8), with $Q = 0.1550$ for several runs, which indicates that the number of edges within groups exceeds the number expected on the basis of chance. These patterns essentially descirbe the co-occurrences and interconnectivity of the binary variables in the corpus. Moreover, since frames consist of such patterns, the three communities indicate that three dominant frames exist in the population and family planning issue.

Scrutinizing the extracted communities of variables reveals a solid framing rationale. In addition, all partitions have at least one representative variable for all frame elements, which have been enumerated in Entman's definition of frame (topic, actor, benefit or risk, solution, and proponent), which means that the key ingredients in defining a frame are in each of the three communities. After a careful investigation of the collection of variables in each community, experts were able to define these three frames: the *population growth and development* frame (Community 1, C1), the *reproductive health* frame (Community 2, C2), and the *family planning/population management threatens moral values* frame (Community 3, C3). A more comprehensive argument regarding how these frames were defined are discussed elsewhere (David, Atun, Legara, & Monterola, 2011). It is important to highlight that the result here is in high agreement with the result obtained by (Legara et al., 2010) using syntactic network analysis where three

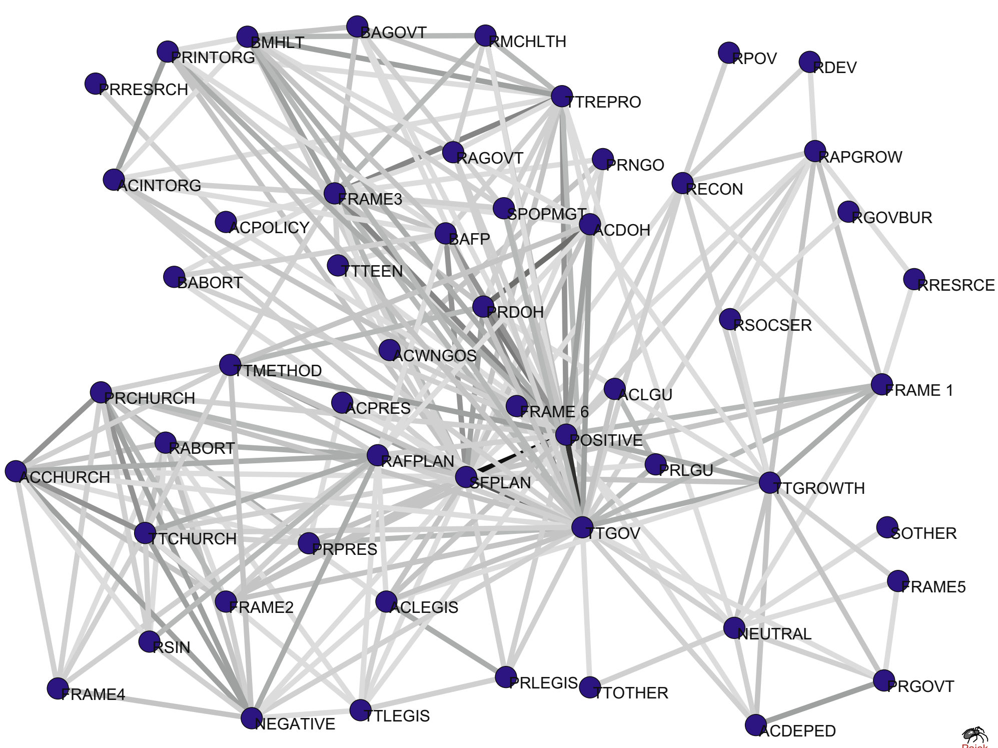

Fig. 6. Nodes in cliques associated with the different frame variables. Cliques were obtained from a reduced network, removing edges e_{kl} with weights $w_{kl} < w_{\text{cutoff}}$, where $w_{\text{cutoff}} = 20$.

Table 9
Comparison of frames obtained using various methods and approaches.

Syntactic network (Legara et al., 2010)	Singular holistic approach	Community detection
k1: Development frame	**F1**: Population and development **F5**: Population growth and demographic trends	**C1**: Population growth and development frame
k2: Maternal health frame	**F3**: Women's and reproductive health	**C2**: Reproductive health frame
k3: Framing by the catholic church	**F2**: Family planning as conflict between government and church **F4**: Population management threatens morals and values	**C3**: family planning/population management threatens moral values frame

Table 10
Schematic coding of articles where $v_1, v_2, \ldots, v_7$ represent sample content analytic variables and `frame1`, `frame2`, ..., `frame6` represent the additional six categorical variables. The values are rated 1 or 0. Moreover, the categorical variables are mutually exclusive; that is, if one of the variables has already been rated as 1.0, the rest would be 0.

	v_1	v_2	v_3	...	v_7	frame1	frame2	...	frame6
article α_1	1	0	1	...	0	1	0	...	0
article α_2	0	0	0	...	0	0	0	...	1
article α_3	1	1	1	...	0	0	0	...	0

frame themes were identified: Development Frame, Maternal Health Frame, and Framing by the Catholic Church (Legara et al., 2010). Table 9 shows a side-by-side tablet comparison of all three differing methods.

5.1. Clique analysis of frame variables

In this section, we compare the validity and reliability of the results obtained using the two methods singular holistic approach (SHA) in Section 3 and community detection approach (CDA) in Section 4. In Section 3, the SHA resulted to six (6) pre-defined frames: the *population and development* frame (F1), the *family planning as conflict between government and church* frame (F2), the *women's and reproductive health* frame (F3), the *population management threatens morals and values* frame (F4), the *population growth and demographic trends* frame (F5), and *others* while in the community detection approach, the computer-assisted algorithm returned three: the *population growth and development* frame (C1), the *reproductive health* frame (C2), and the *family planning/population management threatens moral values* frame (C3).

To quantify the extent of difference and/or similarity of the results, a clique detection method was employed on the corpus together with the SHA results. To do so, we utilized network $G(V, E)$ and added six more nodes or vertices in the network that represent the six categorical variables (`frame1`, `frame2`, ..., `frame6`). Consider that on top of filling-out the "code sheets", the coders were also asked to categorize each article α_i to a dominant frame based on their understanding of the subject matter after a perusal of the text. Including the six categorical variables (CV), a "code sheet" now looks like Table 10. The method for adding the edges connected to the joined categorical variables is the same as in the procedure discussed above. When an article α_i is rated 1.0 in a categorical variable, then that variable has an edge that connects it to all other CAVs that are rated 1.0 in α_i. Needless to say, the categorical variables are mutually exclusive and that the final edges are weighted.

All sets of maximal cliques containing a given CV were then extracted. From a sociological perspective, a *clique* is a small group of individuals with shared interests, patterns of behavior, or other attributes or qualities in common (Jones & Gerard, 1967). In the same light, by calculating all cliques linked to the mutually exclusive CVs, we get to extract other variables (the CAVs) that are "similar" to them. It is conjectured that this commonality or "shared interest" among content analytic variables is associative to the frames that they describe, respectively. Since we are only focused on relationships between variables, we introduced a cutoff edge-weight $w_{\text{cutoff}} = 20$, which is approximately 5% of the total number of articles, whereby all $w_{kl} < w_{\text{cutoff}}$, e_{kl} is removed from $E(G)$. What this cutoff means is that we only consider co-occurrence of two variables k and l if and only if they have been rated 1.0 together in at least 20 of the news articles. This gives us a more reliable and stronger vertex–vertex connection. The stronger the links between the variables are, the more reliable the relationships and therefore, the patterns. Without loss of generality, this process just increases the resolution of the analysis. The resulting network (with both the content analytic variables and the categorical ones) is shown in Fig. 6. For a much easier visual representation of the resulting network, Fig. 7 shows a Venn diagram of the "frame sets".

The obtained cliques in this reduced network are quite distinct albeit there are certain content analytic variables that belong to more than one set (see Fig. 7). Clearly, the frames are not mutually exclusive (due to the overlaps) to each other. This result is rather intuitive.

Increasing the cut-off variable further to 25, $w_{\text{cutoff}} = 25$, more distinct groupings (lesser overlaps) appear (see Fig. 8). The method used also showed that a few variables (`ttgov`, `positive`, `sfplan`) are present in all the cliques identified. Since these elements do not seem to contribute in the uniqueness of frames, we removed them in the subsequent analysis, including the categorical variable `frame6`.

The clique detection method reveals that although there are six (6) predefined frames, three of these are subsets of at least one of the other frame sets. In fact, results show that there are three dom-

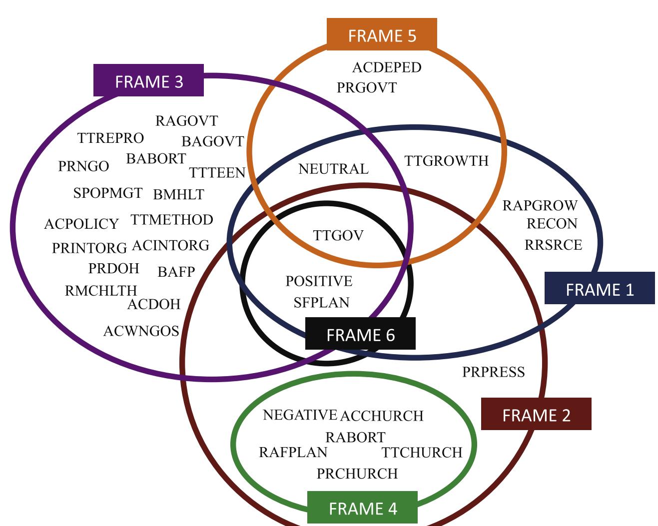

Fig. 7. Venn diagram of the frame sets of the reduced network in Fig. 6 (weights $w_{kl} < w_{\text{cutoff}}$, where $w_{\text{cutoff}} = 20$).

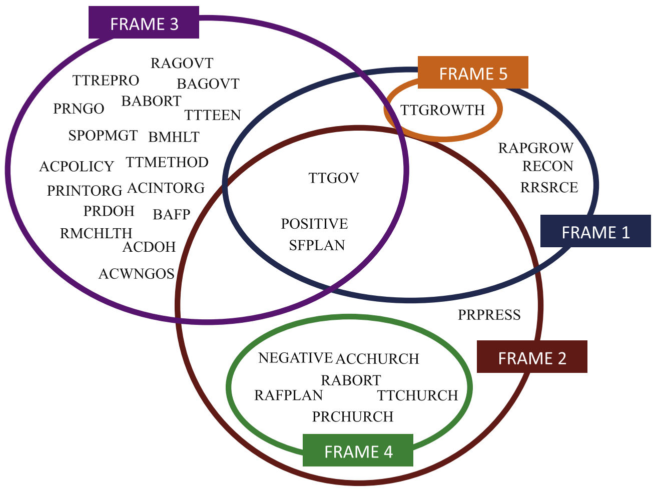

Fig. 8. Venn diagram of the frame sets of the reduced network where edges with weights $w_{kl} < w_{\mathrm{cutoff}}$ were removed, and where $w_{\mathrm{cutoff}} = 20$.

inant sets that are mutually exclusive of each other when the variables `ttgov`, `positive`, `sfplan` are removed: F1, F2, and F3. Moreover, F5 ⊂ F1, F4 ⊂ F2 and finally F6 ⊂ F1, F2, and F3. By simplifying the six pre-defined frame sets to three exclusive ones, results show that the two methods are now in perfect agreement. It is also important to highlight that the results obtained here agree very well with the results recently published that enumerated framing themes surrounding the population issue using syntactic network analysis (Legara et al., 2010).

In conclusion, we performed framing analysis on a corpus of news texts on the population and family planning issue in the Philippines using two distinct approaches. A more traditional singular holistic procedure was initially implemented where coders classified each news text to a specific pre-defined frame. Then, we demonstrated a new framing approach that casts the resulting codings as a network of content analytic variables (CAVs). Here, frames were treated as a collection of elements forming patterns in the CAV network. We showed that both procedures are consistent with the results derived using syntactic network analysis (Legara et al., 2010). Finally, we note that the method illustrated here can be utilized to address the clustering issue that plagued framing scholars in their quantitative exploration of news frames in texts.

Acknowledgments

The authors acknowledge help of J.M. Atun in the data collection and coding process. C.M. and E.F.L are supported by Singapore A*STAR SERC Complex Systems Programme's research grant (1224504056).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov & F. Csake (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.
Bardhan, N. (2001). Translational AIDS–HIV news narratives: A critical exploration of overarching frames. *Mass Communication and Society, 4*(3), 283–309.
Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal Statistical Mechanics: Theory and Experiment*, P10008.
Central Intelligence Agency Online Factbook. ISSN 1553-8133. Website: <https://www.cia.gov/library/publications/the-world-factbook/geos/rp.html>. Last accessed on: December 06, 2012.
David, C., Atun, J. M., Legara, E. F., & Monterola, C. (2011). Finding frames: Comparing two methods of frame analysis. *Communication Methods and Measures, 5*(4), 329–351.
Entman, R. M. (1991). Framing U.S. coverage of international news: Contrasts in narratives of the KAL and Iran air incidents. *Journal of Communication, 41*(4), 6–27.
Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*, 51–58.
Forster, M. R. (2000). Key concepts in model selection: Performance and generalizability. *Journal of Mathematical Psychology, 44*, 205–231.
Gamson, W. A. (1989). News as framing: Comment on Graber. *American Behavioral Scientist, 33*(2), 157–161.
Hoerl, K. E., Cloud, D. L., & Jarvis, S. E. (2009). Deranged loners and demented outsiders? Therapeutic news frames of presidential assassination attempts, 1973-2001. *Communication, Culture & Critique, 2*(1), 83–109.
Joachim, J. (2003). Framing issues and seizing opportunities: The UN, NGOs, and womens rights. *International Studies Quarterly, 47*, 247–274.
Jones, E. E., & Gerard, H. B. (1967). *Foundations of Social Psychology*. New York: Wiley Books.
Koch, J. W. (1998). Political rhetoric and political persuasion: The changing structure of citizens preferences on health insurance during policy debates. *Public Opinion Quarterly, 62*(1), 209–229.
Koella, J. C., (2003). The impact of interest groups and news media framing on public opinion: The rise and fall of the Clinton health care plan. In *Annual Meeting of the Midwest Association for Public Opinion Research. Chicago, IL.*
Kohring, M., & Matthes, J. (2002). The face(t)s of biotech in the nineties: How the German press framed modern biotechnology. *Public Understanding of Science, 11*, 143–154.
Lazarsfeld, P. F. (1950). The logical and mathematical foundations of latent structure analysis. In Samuel A. Stouffer (Ed.), *Measurement and prediction* (pp. 362–412). New York: John Wiley & Sons.
Legara, E., Monterola, C., David, C., & Atun, J. M. (2010). News framing of population and family planning issues via syntactic network analysis. *International Journal of Modern Physics C, 21*(1), 51–65.
Lewis, J., & Linzer, D. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software, 42*(10), 1–29.
Lin, T. H., & Dayton, C. M. (1997). Model selection information criteria for non-nested latent class models. *Journal of Educational and Behavioral Statistics, 22*(3), 249–264.
Matthes, J., & Kohring, M. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication, 58*(2), 258–279.
Miller, M. M. (1997). Frame mapping and analysis of news coverage of contentious issues. *Social Science Computer Review, 15*, 367–368.
Miller, M. M., Andsager, J., & Riechert, B. P. (1998). Framing the candidates in presidential primaries: Issues and images in press releases and news coverage. *Journalism and Mass Communication Quarterly, 75*, 312–324.
Murphy, P. (2001). Affiliation bias and expert disagreement in framing the nicotine addiction debate. *Science, Technology, & Human Values, 26*, 278–299.
Newman, M. (2006). Modularity and community structure in networks. *Proceedings of the National academy of Sciences of the United States of America, 103*(23), 8577–8582.
Newman, M., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*, 026113.
Palla, G., Dernyi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature, 435*, 814–818.
Personal communication with Jörg Matthes.
Roxas, R., & Tapang, G. (2010). Prose and poetry classification and boundary detection using word adjacency network analysis. *International Journal of Modern Physics C, 21*(4), 503–512.
Ryan, C., Caragee, K. M., & Meinhofer, W. (2001). Theory into practice: Framing, the news media, and collective action. *Journal of Broadcasting and Electronic Media, 45*(1), 175–182.
Scheufele, B. T., & Scheufele, D. A. (2010). Of spreading activation, applicability and schemas: Conceptual distinctions and their operational implications for measuring frames and framing effects. In *Doing News Framing Analysis: Empirical and theoretical perspectives*. NY: Routledge.
Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.
Shah, D., Watts, M. D., Domke, D., & Fan, D. P. (2002). News framing and cueing of issue regimes: Explaining Clinton's public approval in spite of scandal. *Public Opinion Quarterly, 66*(3), 339–370.
Sinatra, R., Condorelli, D., & Latora, V. (2010). Networks of motifs from sequences of symbols. *Physical Review Letters, 105*, 178702.
Van Gorp, B. (2005). Where is the frame? Victims and intruders in the Belgian press coverage of the asylum issue. *European Journal of Communication, 20*, 485–508.
Yeo Park & Arabi (2007). News framing west nile virus and outbreak of new health hazard. *Journal of Humanities and Social Sciences, 1*, 1–14.