

Department of Social Systems and Management

Discussion Paper Series

No. 1260

**CORE AND PERIPHERAL SEGMENTS OF WEB FRONTAL  
PAGES IDENTIFIED BY NETWORK ANALYSIS OF  
EYE-TRACKING RECORDS**

by

**MATSDUA, Noriyuki<sup>1</sup> and TAKEUCHI, Haruhiko<sup>2</sup>**

**April, 2010**

<sup>1</sup>Dept. of Social Systems & Management

Grad. School of Systems & Information Engineering, Univ. of Tsukuba

mazda@sk.tsukuba.ac.jp

<sup>2</sup>National Inst. of Advanced Industrial Science & Technology (AIST), Tsukuba

takeuchi.h@aist.go.jp

*(Please contact the first author before quoting the paper.)*

UNIVERSITY OF TSUKUBA

Tsukuba, Ibaraki 305-8573

JAPAN

DISCUSSION PAPER SERIES NO. 1260  
DEPT. OF SOCIAL SYSTEMS AND MANAGEMENT, UNIV. OF TSUKUBA

**CORE AND PERIPHERAL SEGMENTS OF WEB FRONTAL  
PAGES IDENTIFIED BY NETWORK ANALYSIS OF  
EYE-TRACKING RECORDS**

*Noriyuki Matsuda<sup>1</sup> and Haruhiko Takeuchi<sup>2</sup>*

<sup>1</sup>Dept. of Social Systems & Management  
Grad. Schl. of Systems & Information Engineering, Univ. of Tsukuba  
mazda@sk.tsukuba.ac.jp

<sup>2</sup>National Inst. of Advanced Industrial Science & Technology (AIST), Tsukuba  
takeuchi.h@aist.go.jp

**Abstract.** Transitional relationships in the eye-tracking records have not received due attention in the past studies. The present work applied network analysis in an attempt to identify central segments and functional communities around them. The data were obtained from 20 subjects who read 10 frontal Web pages that were classified into three types of layouts. The network, built for each page, represented the transition of fixations among the segments of a 5x5 mesh imposed on the screen. The core and peripheral nodes were identified by multiple centrality indices as well as the ranking scores and they corresponded fairly well to physical locations of the screen. The clique-based communities revealed interesting patterns that ran counter to "banner-blind" and that indicated the effects of the three layout types. We presently plan to incorporate a pattern mining technique(s) in hopes of enhancing the present approach.

(The revised version titled "Networks Emerging from Shifts of Interest in Eye-Tracking Records" is to appear in *eMinds*, 2011, Vol.2(7).)

## **Introduction**

When browsing books at a bookstore, many people tend to guess the quality of an entire book by its cover as is cautioned by a well-known aphorism "Don't judge a book by its cover. The same seems to hold for the web sites of businesses oriented toward consumers (often abbreviated as B2C). Bold as it may be, this belief led us to examine eye movements on the Web's top pages under the widely shared assumption in the field that eye fixations and saccades indicate viewers' interests or attention (Granka et al. [1]; Josephson [2]).

Researchers normally track eye movements across areas set out on a page instead of particular textual or graphical components due to the technology that is available today. One can set out segments by imposing a grid or mesh varying in size to cover different kinds of information such as a navigation menu, the main content, and advertisements (Habuchi et al. [3]; Pan et al. [4]). Alternately, one can set out segments equal in size when, for instance, comparing various pages with different layouts like those in the present work. Segments that attract a great deal of attention are called hotspots, as illustrated in Figure 1-a, which has a frontal web page covered by a 4 x 3 mesh. The size of the circles expresses the amount of fixations by an individual or collective user(s), indicating its hotness: the larger, the hotter (Cutrell and Guan [5]). Thus, one can infer the informativeness or attractiveness of the segments from the hotness. Yet, the analysis of hotness alone is rather insufficient because it is static, and neglects dynamic features, i.e., shifts of interest.

Of the two dynamic approaches (see Figure 1-b and c), scanpath analysis of saccade-fixate-saccade sequences (Goldberg [6]) should help reveal the properties of

individual trajectories, such as their length, smoothness, and spatial density. However, they are hard to synthesize over repeated measures or across individuals. Network analysis, on the other hand, can be applied to either individual or aggregate records, treating hotspots as nodes and the transitions between them as links. By focusing on the immediate relations among nodes, one naturally masks the longer sequential relations along the trajectories that deserve separate treatment as will be explained later in the Discussion. Yet, the advantages of network analysis become clear when it is aimed at locating the center or core of attention and the nearby segments visited by viewers.

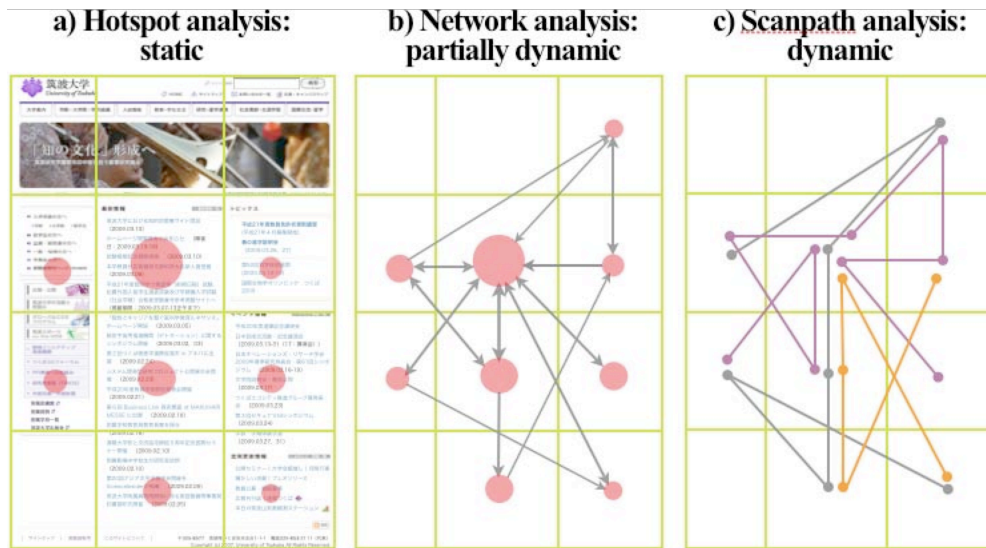


Figure 1: Static and dynamic approaches

However, how should we identify a core node(s) of a network when there is no direct measure of the core-ness in network analysis? We need to derive it from available indices. In this connection, the congruence principle used by Matsuda et al. [7, 8] as well as that by Matsuda and Takeuchi [9] appears to be viable. In their analyses of relatively large undirected networks constructed from the textual information of two B2C web sites, they identified the core nodes on the basis of congruence among multiple centrality indices and importance rankings. In brief, nodes with high centrality and high rankings were judged to be cores. Moreover, they examined the vicinity networks of cores constructed from the neighborhoods of individual cores by union and intersection operations in terms of broad and narrow senses of vicinity.

Stimulated by their approach at a general level, however, we found that some modifications were necessary for the present analysis, taking into consideration the small network size and the directionality of links as well as the presence of tightly connected sub-networks, technically known as *cliques*. In addition, peripheral nodes that consistently failed to attract attention seemed to be worth examining, since such segments were perhaps unwanted on an *e*-market page, whereas they may help avoid the jammed appearance on a fashionable page. It suffices to say that the peripheral nodes can be identified by reversal congruence, i.e., consistently low centrality and low rankings.

It should be of great value to both eye-movement researchers and Web-designers to learn the correspondence between the core and peripheral nodes, on the one hand, and the segments placed on the screen, on the other.

Before closing this section, we will provide a concise explanation of graph/network analysis. Interested readers should also see the Appendix for the glossary. Graphs, in the strictly mathematical sense, consist of points, often called nodes or vertices, and the connections between them, called arcs, edges, or links. Of the various types of graphs are networks and trees (Text Encoding Initiative [10]). In practice, graphs and networks are often used interchangeably and we have followed suit.

Nodes representing sources or origins are linked to the respective targets or destinations in a directed network. No such distinction is made in an undirected network. A *clique* is a tight subgraph in which every pair of nodes is linked. Note that a clique may contain sub-cliques. A *community* in network analysis represents a similar notion in which nodes are relatively densely connected compared to the nodes outside of it. Community detection has been one of the hottest issues in network analysis (see Newman [11]; Newman and Girvan [12]; Raghavan et al. [13]). Put briefly, the directionality led us to incorporate additional ranking scores, and the tight subgraphs led us to form clique-based communities.

The centrality of a node in a given network can be measured by a) the number adjacent nodes—*degree*, b) the average length of the shortest paths from it to the rest—*closeness*, and, c) the proportion of shortest paths in all the pairs of nodes running through it—*betweenness* (See Freeman [14]). Beside these classical indices, the centrality, or importance, can be inferred from the ranking scores of Brin and Page’s [15] *PageRank* and Kleinberg’s [16] *authority*-and *hub*-scores.

## Method

*Subjects*—Twenty residents, (7 males and 13 females), living near a research institute (AIST) located in Japan were recruited for the experiments. They had normal or corrected vision, and their ages ranged from 19 to 48.

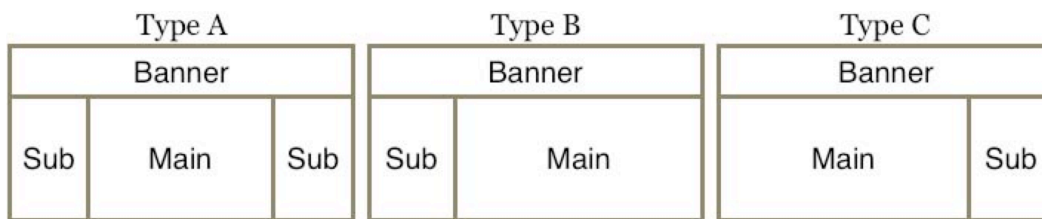


Figure 2: Layout types

*Stimuli*—The frontal pages of ten commercial web sites were selected from various business areas—airlines, e-commerce, finance, and banking. The pages were classified into Types A, B, and C, which differed in their layout of the principal layer, while they all had a banner in the top layer, as shown in Figure 2. The main areas of Type A were sandwiched between sub-areas, while the main areas of Types B and C were accompanied by a single sub-area either on the left (B) or the right (C). The top pages will be referred to, hereinafter, as  $TP_n$ , where  $n$  denotes the page number.

*Apparatus and Procedure*—The stimuli were presented with 1024 x 768 pixel resolution on a 17” TFT display of a Tobii 1750 eye-tracking system with the rate set at 50Hz. The subjects were instructed to look at each page carefully for 20 sec.

*Mesh and Fixation*—A 5x5 mesh was imposed on the effective part of the screen stripped of white margins, which had no text or graphics. The tracking records of each subject were transformed to fixation data on condition that the eyes stayed in the same area for a 100-msec.



Figure 3. An example of 5 x 5 mesh

period. The segments of the mesh were sequentially coded by using a combination of alphabetical labels for the rows and numerical labels for the columns: A1, A2, . . . , A5 for the first row; B1, . . . , B5 for the second; . . . ; and, E1 , . . . , E5 for the fifth.

*Transition matrix*—For network analysis, a square transition matrix was prepared for each page to record the frequencies of fixation shifts from one segment to another aggregated across subjects. Its rows (and columns) were arranged corresponding to the segment codes sorted as [A1, A2, . . . , A5, B1, . . . , B5, . . . , E1, . . . , E5]. The codes were used as the names of network nodes for the ease of inspection.

The diagonal cells of the 25 x 25 matrices contained counts of prolonged fixations, i.e., loops in the network terminology. Finally, the entries of the transition matrices were divided by the respective total frequencies to express the relative volumes. The matrices served as adjacency matrices, the cell values of which were used as weights of links. All the computations and graph layouts were carried by the statistical package R and its library called igraph [17].

## Results

Of the top 10 pages used as stimuli, TP5 was eliminated due to its broad white space. TP4 was also eliminated because of its absence of a solid core, which will be explained later in this section. The remaining pages were subjected to analysis, i.e., TP1, 3, 6, and 8 (Type A), TP2 and 9 (Type B), and TP7 and 10 (Type C).

### 3.1 Basic Properties

The constructed networks were identical in size, and all were comprised of 25 nodes. The loops accounted substantial proportions of link weights—34.7 to 44.2% (Type A), 39.2 to 41.3% (Type B), and 40.4 to 40.9% (Type C). For the sake of simplicity, however, they were omitted from subsequent analysis, which caused no side effects.

As listed in Table 1, the initial networks did not greatly differ from one another on any indices. Yet, an interesting contrast was found in *reciprocity* defined as the ratio of the number of two-way links to the combined number of two-and one-way links. While the one-way links generally outnumbered the two-way links (*reciprocity* < .486), only on TP6 the trend was slightly reversed (.507). High *transitivity*, ranged from .563 to .617, and mild *density* (< .390) indicated the presence of cliques defined as complete subgraphs.

Table 1: Basic properties of initial networks by Type

Index	Type A				Type B		Type C	
	TP1	TP3	TP6	TP8	TP2	TP9	TP7	TP10
Number of links								
two-way	76	58	70	65	63	68	70	65
one-way	82	75	68	81	85	80	74	74
Total $N^a$	234	191	208	211	211	216	214	204
<i>Reciprocity</i>	.481	.436	.507	.445	.426	.459	.486	.468
<i>Transitivity</i>	.617	.582	.593	.598	.609	.568	.575	.563
<i>Density</i>	.390	.318	.347	.352	.352	.360	.357	.340

Note<sup>a</sup>: A two-way link consists of two links in opposite directions, i.e., Total  $N = 2*(N \text{ of two-way}) + (N \text{ of one-way links})$

### 3.2 Identification of core and peripheral nodes

The nodes that were ranked highest (or lowest) at least on three indices were selected as core (or peripheral). There were no ties involved in the rankings. Of the cores listed in Table 2, perfect congruence was only observed on TP1 followed by less perfect ones on TP3, 7, and 10 that were ranked highest on four indices. The rest (TP6, 8, 2, and 9) minimally satisfied the criterion. Interestingly, all the rankings of the cores were highest in *degree* and *closeness* with two mild exceptions (TP8 and 2), probably due to the

node-centric property of the indices, as is explained in the Appendix. The peripheral-ness was more thorough compared to the core-ness in terms of the ranking values (see Table 3). The nodes recorded the lowest or second lowest rankings (i.e., 24 or 25) on the indices with exceptions for TP2 and TP7. Perfect congruence was observed on TP6 and TP9 followed by nearly-perfect ones on TP1, 3, 8 (all of which belonged to Type A), and 10.

Table 2: Core nodes ranked highest on at least three indices by Type

Index	Type A				Type B		Type C	
	TP1	TP3	TP6	TP8	TP2	TP9	TP7	TP10
	B2	D3	C3	C3	C2	B3	B5	B1
<i>Degree</i>	1	1	1	2	1	1	1	1
<i>Betweenness</i>	1	1	2	2	1	3	1	7
<i>Closeness</i>	1	1	1	1	2	1	1	1
<i>PageRank</i>	1	1	5	1	8	4	3	1
<i>Authority-score</i>	1	2	1	9	1	2	1	1
<i>Hub-score</i>	1	6	2	1	5	1	2	8

Table 3: Peripheral nodes ranked lowest on at least three indices by Type

Index	Type A				Type B		Type C	
	TP1	TP3	TP6	TP8	TP2	TP9	TP7	TP10
	E4	E5	E4	E3	E5	D5	E3	E2
<i>Degree</i>	25	25	25	25	25	25	25	25
<i>Closeness</i>	25	24	25	25	23	25	21	25
<i>PageRank</i>	25	25	25	25	24	25	21	24
<i>Authority-score</i>	25	25	25	25	25	25	25	25
<i>Hub-score</i>	25	25	25	24	25	21	25	25

Concerning the correspondence between the computational (i.e., networks) and physical locations, three tendencies were immediately noticeable, as shown in Figure 4. The cores belonged to non-marginal rows B, C, and D, the peripherals belonged to bottom row E or to marginal column 5, and none of them appeared in the top row. The correspondence could have been perfect, had the cores B1 and B5 of TP7 and TP10 (Type C) been located in internal columns.

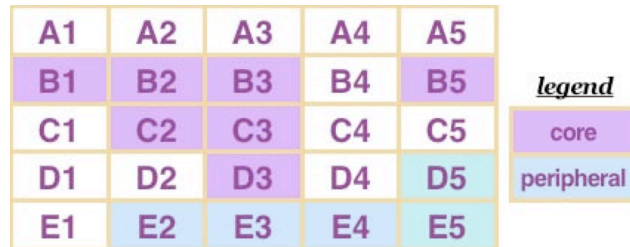


Figure 4: Locations of cores and peripherals on screen

### 3.3 Extraction of clique-based communities

In view of the embedding structure of cliques, only the largest ones were obtained for each network. Their sizes were nearly identical: 7 (TP1, 3, and 6 of Type A, TP9 of B, and TP10 of C) and 8 (TP8 of Type A, TP2 of B, and TP7 of C). However, their numbers varied from 1 (TP8) to 17 (TP3) with 6 on average (see Table 4). Some of the cliques

pertaining to TP1, 3, and 6 of Type A did not contain the cores identified in the entire networks. The missing ratios (the number of cliques without cores to the total number of cliques) were 2/11 (TP1), 10/17 (TP3) and 3/5 (TP5).

Table 4: Properties of cliques and clique-based communities by Type

Index	Type A				Type B		Type C	
	TP1	TP3	TP6	TP8	TP2	TP9	TP7	TP10
Maximal cliques								
<i>n</i> of cliques	11	17	5	1	2	7	3	2
<i>n</i> of cliques <sup>a</sup>	2	10	3	0	0	0	0	0
Clique-based community								
<i>Size</i>	22	14	10	8	10	13	11	8
<i>Reciprocity</i>	.478	.549	.625	.607	.537	.600	.653	.630
<i>Transitivity</i>	.647	.812	.895	1.00	.920	.795	.898	.962
<i>Density</i>	.429	.604	.722	.804	.700	.615	.736	.786

Note<sup>a</sup>: Cliques without cores.

The presence of multiple cliques in all networks except TP8 led us to form a community by union,  $\cup$ , of the clique(s) for each TP. The community size as in the index of the heterogeneity of cliques within individual networks in this study varied from 8 (TP8 and 10) to 22 (TP1) with 12 on average. TP1 indeed yielded a community comprised of almost all nodes of the original network (22 out of 25).

The extracted communities are shown in Figures 5 and 6 in large nodes linked by thick lines over the initial networks. The links connected to the cores are emphasized in different colors according to their directionality, i.e., two-way, inward, and outward.

The community of TP1 was exceptional also with respect to *reciprocity*, *transitivity*, and *density*, probably due to the extreme heterogeneity of the *cliques*. Its reciprocity fell below .500, indicating the relative dominance of one-way links, whereas the opposite was true for all the other communities ( $\geq .537$ ). Interestingly, one-way links were generally dominant in the initial networks (see Table 1). While the *transitivity* of the communities all increased from the initial networks, TP1 recorded the least increase, from .617 to .647, resulting in a sharp decline from the largest to the smallest values. The same held for *density*.

Finally, the communities differed in the inclusion of segments in the marginal rows, A and E. Concerning TP1 whose community size was nearly maximal (22/25), most marginal segments other than A2 and E4 remained in the community. All other communities but that of TP10 (Type C) included some of or all the segments of row A. Particularly noteworthy was the complete inclusion of the top row, [A1, . . . , A5], in the communities of TP9 (Type B) and TP7 (Type C). Furthermore, A1 and some other top segments were mutually linked to the cores (TP1 and 6 of Type A, TP2 and 9 of Type B, and TP7 of Type C). In sharp contrast to the top row, only the communities of TP1 and TP9 contained the marginal segment(s) in row E. Of these, TP9 contained E4 alone.

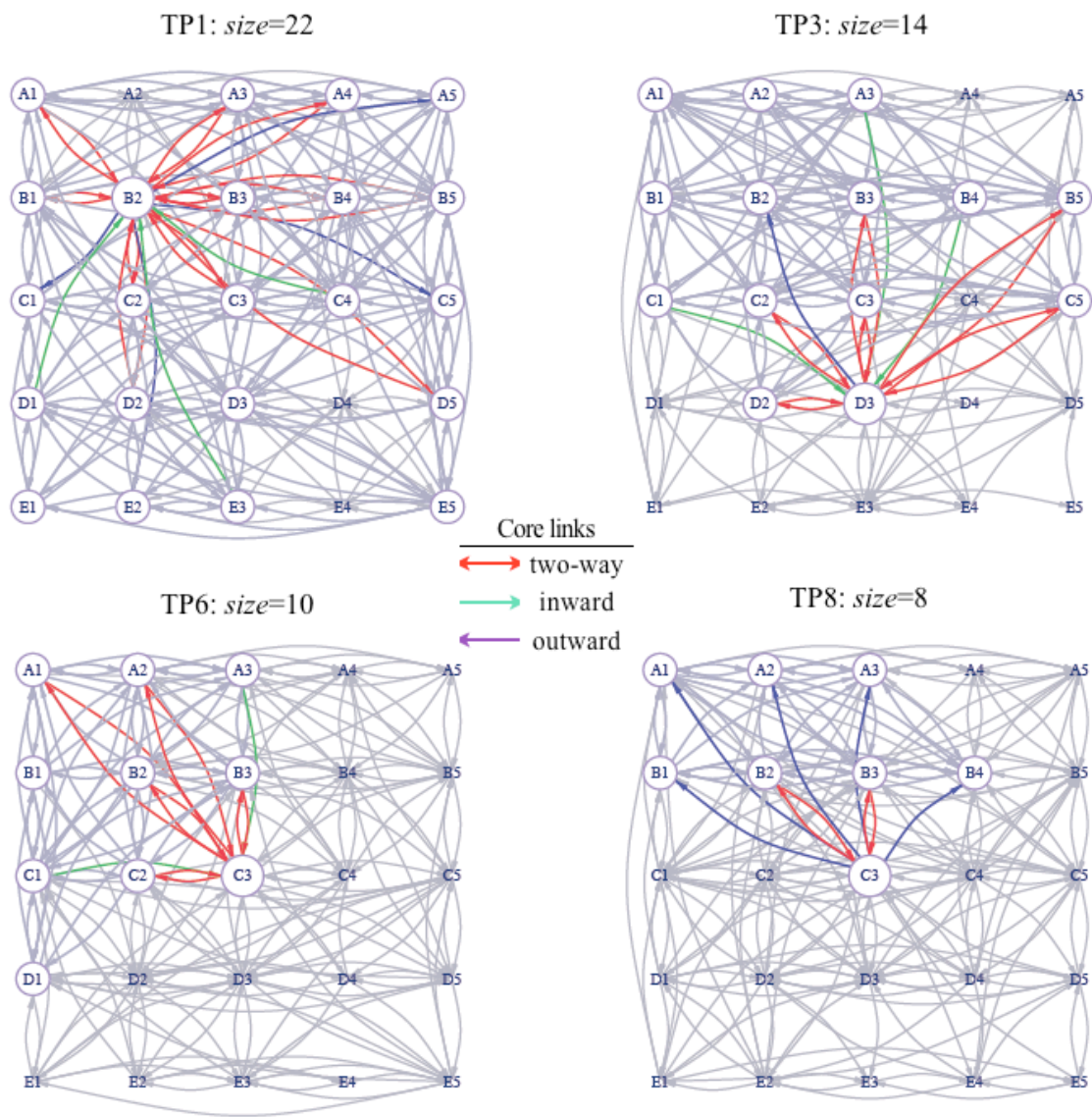


Figure 5: Clique-based communities of Type A



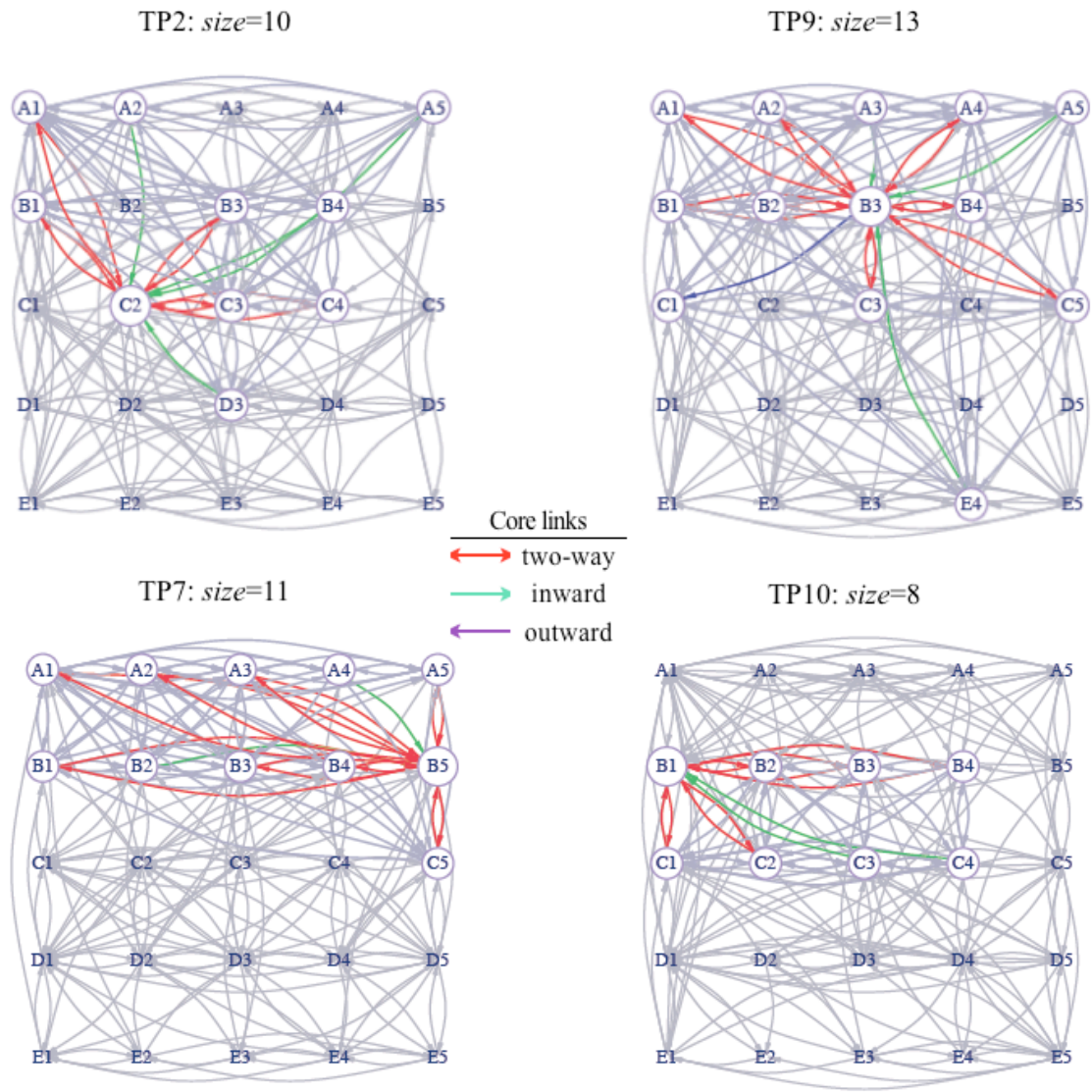


Figure 6: Clique-based communities of Types B (above) and C (below)

## Discussion

The present network analysis revealed how the core and peripheral segments of frontal Web pages could be located from the eye-tracking records of viewers, which were transformed into transition matrices. It deserves special note that the identified cores were truly central not just because of the high rankings on multiple indices, but because of the agreement among local (*degree*) and global (*betweenness* and *closeness*) as well as recursive importance (*PageRank* and *authority- and hub-scores*). By altering the criterion from the high to the low end, we identified peripheral segments as well. Of particular interest were the cores and the peripherals, which corresponded fairly well to physical locations on the screen. Web designers can learn whether their intentions in allocating information were met or not from the results.

The term "banner-blind" (e.g., Pagendam and Schaumburg [18]) refers to the reluctance of people to pay attention to banners often placed at the top of frontal web pages. That no cores represented the segments in row A seems to have confirmed this. On closer examination, however, they were frequently found in the clique-based communities.

Moreover, some of them were found to have mutual links to the respective cores in all types of layouts. Upper-left segment A1 was always present particularly in those cases. Hence, "banner-blind" did not apply to the present records. If it exists, it must be more specific to the content and design rather than being universal. Still, one may argue that this might have been an artifact arising from the omission of loops in the treatment of hotness. We presently have no answers to this. Whether it was an artifact or not certainly deserves fuller investigations in the future.

In contrast to the intermediate importance of top segments, the bottom segments attracted the least amount of interest by viewers. That is, the peripheral segments were concentrated on the bottom row and segment D5. Their locations were remote from the upper left segments. The lack of attention to these segments was in good agreement with the dominance of F-shaped patterns reported by Nielsen [21]

We believe that forming clique-based communities could facilitate the evaluation of web pages, revealing tight linkages between limited number of segments unless the cliques were highly heterogeneous like TP1. With a manageable community size, one could closely examine the linkages of cores to their neighbors, for instance, with respect to the direction and/or weights of the links in light of the intentions in design.

The composition of the cliques and the location of the core segments are indicative of the effect of page layouts. First, the pages of Type A alone produced cliques both with and without the cores. This complexity may be attributed to the presence of sub-areas on both sides of the main area. Second, the cores of Type C pages belonged to marginal columns, while those of Types A and B were located in the internal segments, both horizontally and vertically. The placement of the sub-area on the right of the main area in Type C might have influenced the fixation behavior by subjects. It is well known in brain science that our left and right vision is processed differently. The information presented in the left visual field of both eyes is sent to the right visual cortex, and vice versa. If this were the sole mechanism, we could easily provide design recommendations. However, the right and left hemispheres of the brain (cerebral hemispheres) communicate via the corpus callosum, possibly affecting the initial projections. Further complications arise from the fact that the two eyes do not always converge in reading characters (see Liversedge [19]). Although this is very intriguing, it will require much more evidence from carefully designed studies to fully account for the eye movements of web viewers on a physiological basis.

In lieu of a conclusion, let us mention our current plan, which is to incorporate PrefixSpan (Pei et al. [22]) or any other relevant pattern mining technique(s). By doing this, we may be able to fill in the gaps between the two dynamic approaches—contrasting scanpath and network analyses. That is, to find predominant sub-paths of interest, the entire paths treated in the former are too long and the links connecting pairs of nodes in the latter are too short. A similar idea was proposed by West et al. [20] who demonstrated how the clustering pattern similarities in eye-tracking records were visualized but did not apply network analysis. It is our hope that our plan will lead us to finding clues that will settle the generality of principal scan patterns, i.e., F-shaped (Nielsen et al. [21]), zig-zagged (Lorigo et al. [23], or Z-shaped (as believed by many leaf-let designers).

## References

- [1] Granka, L.A., Hembrooke, H.A. and Feusner, M.K. 2004. Correlates of visual salience and disconnect: An eye-tracking evaluation. Unpublished research report, Cornell University Human-Computer Interaction Lab.

<http://laura.granka.com/research.html>

- [2] Josephson, S. 2004. A summary of eye-movement methodologies. FactOne: [http://www.factone.com/article\\_2.html](http://www.factone.com/article_2.html)
- [3] Habuchi, Y., Takeuchi, H., and Kitajima, M. 2007. Where do web visitors look and at what? –An eye-tracking study. In *Proceedings of HCI 2007*, pp. 1084-1088. April 28-May 3, 2007. San Jose, CA, USA.
- [4] Pan, B., Hembrooke, H., Gay, G., Granka, L., Feusner, M. and Newman, J. 2004. The determinants of web page viewing behavior: An eye tracking study. In *Proceedings of Eye Tracking Research and Applications (ETRA'2004)*, pp. 147-154.
- [5] Cutrell, E. and Guan, Z. 2007. What are you looking for? An eye-tracking study of information usage in web search. In *CHI 2007 Proceedings*, pp. 407-416. April 28-May 3, 2007. San Jose, CA, USA.
- [6] Goldberg, J. H. and Kotval, X. P. 1999. Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, Vol. 24, pp. 631-645.
- [7] Matsuda, N., Machida, A. and Mizuno, K. 2008a. Structural comparisons of the semantic interface of the top cosmetic brands on the web by network analysis. In *Proceedings of IADIS Interfaces and Human Computer Interaction 2008 (IHCI)*, pp. 181-188; Amsterdam, Netherlands, July 20-27, 2008.
- [8] Matsuda, N., Machida, A. and Mizuno, K. 2008b. Visual inspections of the semantic interface of the web: A comparison of the top cosmetic brands by SVD and network. In *Proceedings of IADIS Computer Graphics and Vision 2008 (CGV)*, pp. 282-288; Amsterdam, Netherlands, July 25-27, 2008.
- [9] Matsuda, N. and Takeuchi, H. 2008. Finding core neighborhoods and communities in the web semantics networks by  $\sigma_i u_i$  of SVD–Selection of terms in proportion to  $\sigma_i^2 / \sigma_1^2$ . In *Proceedings of the Japan Behaviormetric Society 2008*, pp. 165-166, Sep.2-5, 2008. Seikei Univ. Tokyo.
- [10] Text Encoding Initiative, The XML version of the TEI guidelines, as of 2009/06/29, <http://www.tei-c.org/Guidelines/P4/html/GD.html>
- [11] Newman, M.E.J. 2006. Modularity and community structure in networks. *PNAS*, June 6, Vol. 103 (23), pp. 8577-8582.
- [12] Newman, M.E.J. and Girvan, M. 2004. Finding and evaluating community structure in networks. *Physical Review*, E 69 (2), 026113.
- [13] Raghavan, U.N., Albert, R., and Kumara, S. 2007. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review*, E 76, 036106, pp. 1-11.
- [14] Freeman, L.C. 1978/79. Centrality in social networks conceptual clarification. *Social Networks*, Vol. 1, pp. 215-239.
- [15] Brin, S. and Page, L. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, Vol. 30, pp. 107-117.
- [16] Kleinberg, J.M. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, Vol. 46 (5), pp. 604-632.
- [17] Csardi, G. 2006. IGraph Library. <http://cneurocv.s.rmk.kfki.hu/igraph/>

- [18] Pagendam, M. and Schaumburg, H. 2001. Why are users banner-blind? The impact of navigation style on the perception of Web banners. *Journal of Digital Information*, Vol. 2 (1), No. 47: Retrieved from <http://journals.tdl.org/jodi/article/view/36/38>
- [19] Liversedge, S.P. 2008. Fixation disparity during reading: Fusion, not suppression. *Journal of Eye Movement Research*, Vol. 2 (3), 5, pp. 1-7.
- [20] West, J.M., Haake, A.R., Evelyn, P., Rozanski, E.P. and Karn, K.S. 2006. eyePatterns: software for identifying patterns and similarities across fixation sequences. In *Proceedings of Eye Tracking Research and Applications (ETRA'2006)*, pp. 149-154.
- [21] Nielsen, J. and Pernice, K. 2009. *Eyetracking Web Usability*. Berkeley, CA: New Riders Press.
- [22] Pei, J., Han, J.B., Mortazavi-Asl, B. and Pinto, H. 2001. PrefixSpan: Mining sequential patterns efficiently by PrefixProjected pattern growth. In *Proceedings 2001 International Conference on Data Engineering (ICDE'01)*, pp. 215-224, Heidelberg, Germany, April 2001.
- [23] Lorigo, L., Haridasan, M., Brynjarsdottir, H., Xia, L., Joachims, T., Gay, G., Granka, L., Pellacini, F., and Pan, B. 2008. Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology*, Vol. 59 (7), pp. 1041-1052.

### Appendix: Glossary

**Clique:** A complete subgraph in which every pair of nodes is connected either unilaterally or bilaterally.

**Density:** The ratio of the number of links,  $n$ , to the maximally possible number, i.e.,  $n/N C_2$ , where  $N$  is the total number of nodes. For a directed network, the denominator is the maximally possible number of two-way links, i.e.,  $2 * N C_2$ .

**Reciprocity:** The ratio of the number of two-way links to the sum of the number of two-and one-way links.

**Centrality indices:** *Degree*, *closeness* and *betweenness* are well-known (see Freeman [14]). The former two are node-centric in that the *degree* of a node is the number of links directly attached to it and the *closeness* of a node is inversely proportional to the average distance (geodesics) to other nodes in a given network. Also based on geodesics  $g$ , *betweenness* uses the information differently with emphasis on the intervening role in the global structure as opposed to the local dominance of degree. The *betweenness* of a node is determined by the extent a node intervenes in transactions or flows between all pairs of nodes. The formulas for *closeness* and *betweenness* of node <sub>$v$</sub>  are:

$$closeness_v = \left( \sum_{i \neq v} g_{iv} / (\text{the number of nodes} - 1) \right)^{-1}$$

$$betweenness_v = \sum_{i \neq v, i \neq j, v \neq j} g_{ivj} / g_{ij}$$

The three indices coincide for a node located in the middle of a wheel-or a star-like network.

**Ranking scores:** *PageRank* (Brin et al. [15]) recursively determines the importance of a node by the importance of the nodes connecting to it, and their importance is further determined by the importance of the nodes connecting to them. Computationally, one first normalizes the adjacency matrix,  $\mathbf{A}$ , by row, and, obtains the scores from the leading eigenvector of its transpose. When *PageRank* is applied to an undirected network, links are treated as bidirectional. *Authority*-and *hub-scores* (Kleinberg [16]) have mutually reinforcing relationships: The authoritativeness of a node is enhanced by the hubness of the nodes linking to it, while the hubness of a node increases as a function of the authoritativeness of the nodes they link to. Interestingly, authority-and hub-scores can be obtained from the leading eigenvectors of  $\mathbf{A}^T\mathbf{A}$  for the former and  $\mathbf{A}\mathbf{A}^T$  for the latter. Thus, they are identical when applied to an undirected network.