Topological Spaces of the Web

Gabriel Ciobanu
Romanian Academy
Institute of Computer Science
Iaşi, Romania
gabriel@iit.tuiasi.ro

Dănuţ Rusu
Faculty of Mathematics
"A.I.Cuza" University
Iaşi, Romania
drusu@uaic.ro

Categories and Subject Descriptors:
Primary Classification: F. Theory of Computation
Additional Classification: G. Mathematics of Computing
General Terms: Theory, Measurement
Keywords: web metrics, topology, density, separation.

1. INTRODUCTION

We study the representation of space over the Web. Web's spatiality is described by (hyper)links, and thus the web space is given by a set of pages and links between them. The complexity of this structure derives not only from the quantity, but also from the extreme dynamics that it presents. This structure is greatly influenced by its dynamics; new documents appear together with their links and disappear, cutting off all the corresponding arcs in the graph. We introduce and study some topologies over web, providing a suitable formal framework to express web connectivity, density, how the whole web space can be generated by a certain subset of pages, and web separation. Following previous papers devoted to time [1] and space [3], we develop a high-level description of spatio-temporal relations to catch the dynamics of the Web documents and the links between them.

2. SPACE MODEL OF THE WEB

The web does not carry the notion of distance in the usual way; a space model of the Web should not be based on the common notion of Euclidean space.

Let us denote by $\mathcal{W}$ the universal set of the Web, as the collection of all pages available on-line, referred through their URIs. We define the binary relation called *points-to*, $\hookrightarrow$: $\mathcal{W} \longrightarrow \mathcal{W}$, by $a \hookrightarrow b$ if document a contains a link to document b. We also define the inverse relation *is-pointed-by*, $\hookleftarrow$: $\mathcal{W} \longrightarrow \mathcal{W}$, $\hookleftarrow = \hookrightarrow^{-1}$. We do not consider internal anchors from a to a as links.

General topology defines and studies some useful properties of spaces and maps, such as connectedness, separation, compactness and continuity. The topologies provide structures that allow to formalize these properties. A space X with a topology τ is called a topological space, and it is denoted by (X, τ). These spaces can be analyzed and described by the topological notions as: neighborhood, closure, open set, closed set, connected set, etc. Over the same space we can define more topologies, and generally we chose the topologies corresponding to the requirements imposed by the modeling process. A topology on the space $\mathcal{W}$ is a family τ of subsets of $\mathcal{W}$ such that: 1. $\emptyset, \mathcal{W} \in \tau$; 2. $\forall \{D_i\}_{i\in I} \subset \tau$, $\cup_{i\in I} D_i \in \tau$; 3. $\forall D_1, D_2 \in \tau$, $D_1 \cap D_2 \in \tau$. The sets of a topology τ are called τ-open sets [5]. A set $A \subset \mathcal{W}$ is τ-closed iff its complement $\mathcal{W} \setminus A$ is τ-open.

Thinking in terms of Web navigation, we define a sailing trip over the Web documents. Let $a, b \in \mathcal{W}$. A *sailing trip* from document a to document b is a function $f : \{1, ..., n\} \rightarrow \mathcal{W}$, where $n \geq 2$ and $a = f(1) \hookrightarrow f(2) \hookrightarrow \ldots \hookrightarrow f(n) = b$. In this case n is called the *length of the sailing trip*, and it is denoted by $\lambda(f)$. The image of f is denoted by $Im(f)$. We can note that $\lambda(f) \geq card(Im(f))$. Let L_{ab} be the set of all sailing trips from document a to document b. Then $\mathcal{L}_{ab}$ is L_{ab} if $a \neq b$, and $\mathcal{L}_{aa}$ is $L_{aa} \cup \{\emptyset\}$. Since we do not consider internal anchors from a to a as links, we have an empty sailing trip $\emptyset$ in $\mathcal{L}_{aa}$ ($\lambda(\emptyset) = 0$).

Let $a, b \in \mathcal{W}$. We say that a is *connected* with b if $\mathcal{L}_{ab} \neq \emptyset$, and this is denoted by $a \rightsquigarrow b$. We say that a is *biconnected* with b if a is connected with b and b is connected with a; this is denoted by $a \leftrightsquigarrow b$.

PROPOSITION 1. *The relation $\rightsquigarrow$ is a quasi-order on $\mathcal{W}$. The relation $\leftrightsquigarrow$ is an equivalence on $\mathcal{W}$.*

Let $a \in \mathcal{W}$, and $A \subseteq \mathcal{W}$. We have the following notations:

$Out_a = \{x \in \mathcal{W} \mid a \rightsquigarrow x\}$, $In_a = \{x \in \mathcal{W} \mid x \rightsquigarrow a\}$,
$Net_a = In_a \cup Out_a = \{x \in \mathcal{W} \mid a \rightsquigarrow x \text{ or } x \rightsquigarrow a\}$,
$R_a = In_a \cap Out_a = \{x \in \mathcal{W} \mid a \leftrightsquigarrow x\}$ – the *center* of Net_a,
$Out_A = \cup_{a\in A} Out_a$, $\quad In_A = \cup_{a\in A} In_a$,
$Net_A = In_A \cup Out_A = \cup_{a\in A} Net_a$, $\quad R_A = \cup_{a\in A} R_a$,
$O_a = \{(x, y) \in \mathcal{W} \times \mathcal{W} \mid a \rightsquigarrow y \Rightarrow a \rightsquigarrow x\}$,
$I_a = \{(x, y) \in \mathcal{W} \times \mathcal{W} \mid y \rightsquigarrow a \Rightarrow x \rightsquigarrow a\}$,
$\triangle = \{(x, x) \mid x \in \mathcal{W}\}$.

PROPOSITION 2. *Let $a, b \in \mathcal{W}$. Then we have*
1. $b \in In_a$ iff $a \in Out_b$;
2. if $b \in In_a$, then $In_b \subseteq In_a$, and $Out_a \subseteq Out_b$;
3. if $b \in Out_a$, then $Out_b \subseteq Out_a$, and $In_a \subseteq In_b$;
4. if $b \in R_a$, then $In_b = In_a, Out_a = Out_b, Net_b = Net_a$;
5. $I_a = (\mathcal{W} \times (\mathcal{W} \setminus In_a)) \cup (In_a \times \mathcal{W})$;
6. $(\mathcal{W} \times \mathcal{W}) \setminus I_a = (\mathcal{W} \setminus In_a) \times In_a$;
7. $O_a = (\mathcal{W} \times (\mathcal{W} \setminus Out_a)) \cup (Out_a \times \mathcal{W})$;
8. $(\mathcal{W} \times \mathcal{W}) \setminus O_a = (\mathcal{W} \setminus Out_a) \times Out_a$;
9. $\triangle \subset O_a \cap I_a$;
10. O_a and I_a are transitive relations.

Let $U \subset \mathcal{W} \times \mathcal{W}$, and $x \in \mathcal{W}$.
We consider the projection $U[x] = \{y \in \mathcal{W} \mid (x, y) \in U\}$.

PROPOSITION 3.
1. $x \rightsquigarrow y \Rightarrow I_a[y] \subseteq I_a[x]$, and $O_a[x] \subseteq O_a[y]$, $\forall a \in \mathcal{W}$.
2. $x \rightsquigarrow y$ iff $y \in I_y[x]$ iff $x \in O_x[y]$.

WWW 2005, May 10–14, 2005, Chiba, Japan.
ACM 1-59593-051-5/05/0005.

We denote by $\preceq_i = \cap_{a\in\mathcal{W}} I_a$, and $\preceq_o = \cap_{a\in\mathcal{W}} O_a$. It is easy to prove that $\preceq_i$ and $\preceq_o$ are quasi-orders on $\mathcal{W}$.

Using these quasi-orders we can define some topologies on $\mathcal{W}$. A topology can be defined directly by means of its open sets, or it can be defined by means of a neighborhood operator. A neighborhood operator is a function $\mathcal{V} : \mathcal{W} \to \mathcal{P}(\mathcal{P}(\mathcal{W}))$ such that $\mathcal{V}(x)$ holds the following, for all $x \in \mathcal{W}$:

1. if $V \in \mathcal{V}(x)$, then $x \in V$;
2. if $V_1, V_2 \in \mathcal{V}(x)$, then $V_1 \cap V_2 \in \mathcal{V}(x)$;
3. if $V \in \mathcal{V}(x)$ and $V \subset U$, then $U \in \mathcal{V}(x)$;
4. $\forall V \in \mathcal{V}(x)$, $\exists W \in \mathcal{V}(x)$ such that $V \in \mathcal{V}(y)$, $\forall y \in W$.

If τ is a topology on $\mathcal{W}$, then $\mathcal{V}_\tau : \mathcal{W} \to \mathcal{P}(\mathcal{P}(\mathcal{W}))$ defined by $\mathcal{V}_\tau(x) = \{V \subset \mathcal{W} \mid \exists D \in \tau \text{ such that } x \in D \subset V\}$ is a neighborhood operator on $\mathcal{W}$. If $\mathcal{V}$ is a neighborhood operator on $\mathcal{W}$, then $\tau_{\mathcal{V}} = \{D \subset \mathcal{W} \mid D \neq \emptyset \text{ and } D \in \mathcal{V}(x), \forall x \in D\} \cup \{\emptyset\}$ is a topology on $\mathcal{W}$. The notions of topology and neighborhood operator are equivalent because $\tau_{\mathcal{V}_\tau} = \tau$ and $\mathcal{V}_{\tau_{\mathcal{V}}} = \mathcal{V}$ [2, 5].

Let $\mathcal{V}_i : \mathcal{W} \to \mathcal{P}(\mathcal{P}(\mathcal{W}))$, $\mathcal{V}_i(x) = \{V \subset \mathcal{W} \mid \preceq_i [x] \subseteq V\}$, and $\mathcal{V}_o : \mathcal{W} \to \mathcal{P}(\mathcal{P}(\mathcal{W}))$, $\mathcal{V}_o(x) = \{V \subset \mathcal{W} \mid \preceq_o [x] \subseteq V\}$.

PROPOSITION 4. *$\mathcal{V}_i$ and $\mathcal{V}_o$ are neighborhood operators defining topologies over $\mathcal{W}$.*

Let $\tau_i = \{D \subset \mathcal{W} \mid D \neq \emptyset \text{ and } D \in \mathcal{V}_i(x), \forall x \in D\} \cup \{\emptyset\}$ be the topology generated by $\mathcal{V}_i$, and $\tau_o = \{D \subset \mathcal{W} \mid D \neq \emptyset \text{ and } D \in \mathcal{V}_o(x), \forall x \in D\} \cup \{\emptyset\}$ be the topology generated by $\mathcal{V}_o$. We call τ_i the *in-topology*, and τ_o the *out-topology*.

If $A \subseteq \mathcal{W}$, then $int_i A$, $cl_i A$, $Fr_i A$ are respectively the interior, the closure, and the frontier of A with respect to τ_i, and $int_o A$, $cl_o A$, $Fr_o A$ are respectively the interior, the closure and the frontier of A with respect to τ_o.

THEOREM 1. *Let $x, y \in \mathcal{W}$.*

1. *If $x \rightsquigarrow y$, then $\mathcal{V}_i(x) \subseteq \mathcal{V}_i(y)$ and $\mathcal{V}_o(y) \subseteq \mathcal{V}_i(x)$.*
2. *$In_a = cl_i\{a\}$, and $Out_a = cl_o\{a\}$, $\forall a \in \mathcal{W}$.*
3. *$x \rightsquigarrow y$ iff $x \preceq_i y$ iff $y \preceq_o x$ ($\preceq_i$ and $\preceq_o$ are dual).*

Consequently, $\preceq_i [x] = Out_x$, and $\preceq_o [x] = In_x$, $\forall x \in \mathcal{W}$. Hence $\preceq_i [x]$ is a τ_o-closed set, and $\preceq_o [x]$ is a τ_i-closed set. Since $cl_i\{a\} = In_a$ and $cl_o\{a\} = Out_a$, $\forall a \in \mathcal{W}$, it follows that, $\forall A \subset \mathcal{W}$, $cl_i A = In_A$, $int_i A = \mathcal{W} \setminus In_{\mathcal{W}\setminus A}$, $Fr_i A = In_A \cap In_{\mathcal{W}\setminus A}$, and also, $cl_o A = Out_A$, $int_o A = \mathcal{W} \setminus Out_{\mathcal{W}\setminus A}$, $Fr_o A = Out_A \cap Out_{\mathcal{W}\setminus A}$. Since $cl_i A = A \cup Fr_i A$, we have $In_A = A \cup (In_A \cap In_{\mathcal{W}\setminus A})$. We have also $cl_o A = Out_A = A \cup (Out_A \cap Out_{\mathcal{W}\setminus A})$. From these remarks we obtain the following characterizations of the closed/open sets.

PROPOSITION 5. *Let $A \subset \mathcal{W}$.*

1. *A is τ_i-closed if only if $In_A \subseteq A$.*
2. *A is τ_i-open if only if $A \cap In_{\mathcal{W}\setminus A} = \emptyset$.*
3. *A is τ_o -closed if only if $Out_A \subseteq A$.*
4. *A is τ_o-open if only if $A \cap Out_{\mathcal{W}\setminus A} = \emptyset$.*

Based on these topological notions, we present some of our results regarding the *connectivity, density* and *separation* of the Web.

Since $\{a\}$ is a τ_i-connected set, it results that $cl_i\{a\}$ is a τ_i-connected set. Therefore In_a is τ_i-connected. In a similar way, since $\{a\}$ is a τ_o-connected set, it results that $cl_o\{a\}$ is a τ_o-connected set, and Out_a is τ_o-connected. Moreover, considering $A \subset \mathcal{W}$, if A is a τ_i-connected set, then In_A is a τ_i-connected set, and if A is a τ_o-connected set, then Out_A is a τ_o-connected set.

THEOREM 2. *If f is a sailing trip from a to b, then $Im(f)$ is a τ_i-connected set, and also a τ_o-connected set.*

A web page $a \in \mathcal{W}$ is called an *α-point* if $In_a \subseteq Out_a$. A web page $a \in \mathcal{W}$ is called an *ω-point* if $Out_a \subseteq In_a$. It is obvious that a is an α-point iff $Net_a = Out_a$, and $a \in \mathcal{W}$ is an ω-point iff $Net_a = In_a$. We denote by Γ the set of α-points, and by Ω the set of all ω-points. Considering $a \in \mathcal{W}$ and $B \subseteq \mathcal{W}$, we call B as an in-branch of a if there are $b \in \Gamma$, and $f \in L_{ba}$ such that $B = Im(f)$. In a similar way, B is an out-branch of a if there are $b \in \Omega$ and $f \in L_{ab}$ such that $B = Im(f)$. Finally, B is a branch of a if B is either an in-branch, or an out-branch of a. Every branch of a is a τ_i-connected set, as well as a τ_o-connected set. Let $\mathcal{B}_{i,a}$ be the set of all in-branches of a, $\mathcal{B}_{o,a}$ the set of all out-branches of a, and $\mathcal{B}_a = \mathcal{B}_{i,a} \cup \mathcal{B}_{o,a}$.

Since for all $a \in \mathcal{W}$, $In_a = \cup_{B\in\mathcal{B}_{i,a}} B$, $Out_a = \cup_{B\in\mathcal{B}_{o,a}} B$, and $Net_a = \cup_{B\in\mathcal{B}_a} B$, then In_a, Out_a and Net_a are τ_i,τ_o-connected sets.

The following result shows that α-points, as well as the ω-points can generate the whole web space with respect to our topologies.

THEOREM 3. *$\mathcal{W}$ is generated by Γ or by Ω:*

1. *Γ is a τ_o-dense set, i.e. $\mathcal{W} = cl_o(\Gamma) = Out_\Gamma$.*
2. *Ω is a τ_i-dense set, i.e., $\mathcal{W} = cl_i(\Omega) = In_\Omega$.*

According to some previous results, we have $\preceq_i \neq \triangle$ and $\preceq_o \neq \triangle$. Therefore $(\mathcal{W}, \tau_i)$ and $(\mathcal{W}, \tau_o)$ are not Hausdorff separate, i.e., they are not T_2 separate (see [5]). Since generally single point subsets of $\mathcal{W}$ are not τ_i (or τ_o) closed, the spaces $(\mathcal{W}, \tau_i)$ and $(\mathcal{W}, \tau_o)$ are not T_1 separate.

We define $d_i(x, y) = card\{a \in \mathcal{W} \mid (x, y) \notin I_a\}$, and $d_o(x, y) = card\{a \in \mathcal{W} \mid (x, y) \notin O_a\}$. Let $x \in \mathcal{W}$ and $\phi_{i,x} : \mathcal{W} \to [0, 1]$, defined by $\phi_{i,x}(y) = \min\{1, d_i(x, y)\}$, $\forall y \in \mathcal{W}$. Because d_i is τ_i-uniformly continuous, $\phi_{i,x}$ is τ_i-continuous. Then we have $\phi_{i,x}^{-1}(\{0\}) = Out_x$ and $\phi_{i,x}^{-1}(\{1\}) = \mathcal{W} \setminus Out_x$. Therefore Out_x is a τ_i-clopen set, i.e. it is an open and closed set simultaneously. Similar, we define the function $\phi_{o,x} : \mathcal{W} \to [0, 1]$, defined by $\phi_{o,x}(y) = \min\{1, d_o(x, y)\}$, $\forall y \in \mathcal{W}$. Then $\phi_{o,x}^{-1}(\{0\}) = In_x$, $\phi_{o,x}^{-1}(\{1\}) = \mathcal{W} \setminus In_x$, and In_x is a τ_o-clopen set. Moreover, $\forall A \subset \mathcal{W}$, In_A is a τ_o-clopen set and Out_A is a τ_i-clopen set.

THEOREM 4. *Let $F \subsetneq \mathcal{W}$ and $x \notin F$.*

1. *If F is a τ_i-closed set, then $F \subseteq \phi_{i,x}^{-1}(\{1\})$.*
2. *If F is a τ_o-closed set, then $F \subseteq \phi_{o,x}^{-1}(\{1\})$.*

COROLLARY 1. *Let $A \subseteq \mathcal{W}$.*

1. *If $x \notin In_A$, there exists $b \in \mathcal{W}$ such that $x \in In_b$, and $In_b \cap In_A = \emptyset$.*
2. *If $x \notin Out_A$, then there exists $b \in \mathcal{W}$ such that $x \in Out_b$, and $Out_b \cap Out_A = \emptyset$.*

3. REFERENCES

[1] Buraga, S., Ciobanu, G.: A RDF-based Model for Expressing Temporal Relations Between Web Sites, *Proceedings WISE'02*, 355-362, IEEE Press (2002).

[2] Cech, E., *Topological Spaces*. Publishing House of the Czechoslovak Academy of Sciences, Prague (1966).

[3] Ciobanu,G., Dumitriu,D.: Space and Time Over the Web (poster), *12th Int'l World Wide Web Conference*, Budapest (2003).

[4] Dhyani, D. Ng, W.K., Bhowmick, S.: A Survey of Web Metrics, *ACM Computing Surveys* vol.34 (4), 469-503 (2002).

[5] Engelking, R., *General Topology*. 2nd edition. Sigma Series in Pure Mathematics 6, Heldermann (1989).