

## Chapter 3

# ANALYSIS AND VISUALIZATION OF SOCIAL DIFFUSION PATTERNS IN THREE-DIMENSIONAL VIRTUAL WORLDS

Shashikant Penumarthy and Katy Börner

### 1. Introduction

In order to indicate how difficult it is to orient oneself and navigate in virtual worlds, let us begin with an example—the story of John. Just having finished an hour of e-mail, John looks at his watch and realizes that it is time to switch identities. A few mouse clicks, the brief appearance of a splash screen and faster than one can say “Avatar!”, the monitor screen fills up and one world is replaced by another.

Exit John, enter PringleCrow. As LinkWorld (one of several three-dimensional virtual worlds that John has access to) fades into view, PringleCrow waits to orient herself (today John has chosen a female avatar). She waits for her world to manifest itself, one object at a time. The network isn't very fast and PringleCrow is suddenly “lost”, the connection to her life-force severed, as John's consciousness comes back to the real world wondering why the graphics are so jerky. Back again to PringleCrow, the life-force restored! This time, though, enough of the world has loaded ensuring that PringleCrow won't be lost again soon. An extension of consciousness takes over, as the human and his avatar are unified. Today, PringleCrow is going to be part of a treasure hunt that is designed to test how well the inhabitants of the virtual world understand the principles of physics. When she teleports to the meeting point for the participants, she realizes that the other avatars are already there, the chatter almost “deafening”. She spots a chat utterance by CyberDog, her team-mate for the treasure hunt, who has been waiting for PringleCrow for a few minutes. A few more chat utterances later (“I am a blue puppy”, “Let's meet near the waterfall”, “It's at coordinates 10 west 56 north I think”), they finally meet. At this point neither avatar is aware of where the rest of the avatars are or

even how many of them are in the world. They walk together to the podium where the avatar StrazyFoure, who is the moderator for the hunt, is reviewing the rules. After the rules are done, StrazyFoure raises his hand to signal the start of the treasure hunt and PringleCrow and CyberDog begin looking for clues. They are careful to communicate through “whispers”, rather than chat to ensure that everything they say is “inaudible” to the avatars from competing teams.

At the end of the day, StrazyFoure goes to each team to find out what they think about the hunt and is a little concerned to discover that not everyone has succeeded to find all answers. While some found the clues very difficult to decipher, others complain that they sometimes could not figure out if they were going the right way. The virtual world is so big that some lost sense of how far they had come from the starting point.

At what point did the unsuccessful avatars get lost? Which clues were easy and which ones were difficult? Did the avatars interpret the clues correctly? What did the avatars talk about during the whole period? Could some of the clues have been placed in locations that were easier to access? These are just a few of the interesting questions that could be asked in order to determine how successful the treasure hunt was and how to support information foraging in 3D virtual worlds.

The avatars of the virtual world as well as researchers that study them need to be able to quickly obtain a broad overview of activities in the virtual world, as well as examine local details. Researchers must also be able to discover patterns of behavior, of movement, chat and interaction on a local and global scale. Some questions are short term such as “What paths did the avatars take?” or “Which avatars’ movement trails deviated most from the average of the entire group?” Other questions are: “Which areas of the world are most used?”, “What topics are the inhabitants talking about and where?”, or “Who talked/interacted with whom?”

In this chapter we describe how advanced data mining and information visualization techniques [1–4] can be advantageously applied to augment, evaluate, optimize, and study collaborative 3D virtual worlds and to study their evolving communities [5]. We begin by discussing related work in the area of analysis and visualization of spatial and social data. After briefly commenting on the nature of virtual world data, we outline types of user groups and their tasks in virtual worlds. The use of patterns for analysis of virtual world behavior is detailed. We then show how techniques developed in areas as diverse as data analysis [6, 7] and information visualization can be combined to create simple yet powerful means for answering some of the questions posed above. A toolkit [8] for the analysis and visualization of social diffusion patterns in virtual worlds is presented. The chapter concludes with directions for future research in analysis and visualization in virtual worlds.

## **2. Related Work**

Several highly diverse research areas have developed theories, techniques and systems for analysis and visualization of data possessing a distinct spatial component. The well-developed area of scientific visualization deals with applying computer graphics to scientific data “for purposes of gaining insight, testing hypothesis, and general elucidation” [9]. Scientific data usually takes the form of measured values obtained from a system or a phenomenon, in addition to three-dimensional coordinates associated with each measurement. Scientific visualization typically consists of data that can be accurately represented using three-dimensional geometry and is being used in such diverse areas as designing aircraft and constructing complete three-dimensional views of the human anatomy [10].

Another mature field that makes extensive use of visualization of spatial patterns is geography. Today’s Geographic Information Systems (GIS) contain a large number of spatial analysis methods that enable one to analyze, model and query spatial data. Cartographers, in particular, have developed a number of techniques to visualize spatio-temporal diffusion patterns. Pioneering work by Dorigo and Tobler [11] represents and visualizes diffusion potentials and gradients as vector fields computed using a continuous spatial gravity model. By overlaying population density information over a geographic map, they have shown that New York, being densely populated, exhibits the highest outward pressure, i.e., it acts as a source, while Florida has the highest inward pressure, i.e., it acts as a sink.

The work that relates most closely to ours comes from the area of social visualization. Social visualizations are a special type of information visualization that focus on analysis of social behavior. For example, lifeline visualizations reveal migrations, transitions and trajectories of users or user groups [12]. Other research aims at the visualization of large-scale conversations, such as those that take place on the Usenet [13] or visualization of access patterns of users on the Web [14]. There also exists work on visualizing and supporting social interactions in text-based or 2D graphical systems. For example, Chat Circles [15] is a 2D graphical interface for synchronous conversation that visualizes the non-textual components of online chatting, such as pauses and turn-taking behavior, which are typically not available in chat log files. Another piece of work analyzes gestures and movement of users in VChat, a graphical chat system [16], in which a comparison is made between average distance and orientation of users in relation to users targeted in their chat and randomly selected users. Naper [17] was among the first to analyze chat text logged in a 3D virtual worlds and called for the use of visual semiotics for the analysis of computer mediated communication.

A very interesting body of work that analyzes social patterns exists in the area of urban studies. Whyte [18] determined the influence of steps, fountains,

green spaces, sitting places, building arrangement, etc. on the crowd flow and social interaction in New York plaza. The company Space Syntax Limited (<http://www.spacesyntax.com/>) undertakes projects that aim to quantify the degree to which urban planning influences socio-economic factors such as pedestrian flow, crime patterns and land value. Results augment strategic design as well as the selection of design alternatives that best serve the needs of a particular segment of population.

Work on mapping MUDs and 3D virtual worlds is a relatively new area. The section on MUDs and Virtual Worlds in Dodge's and Kitchin's "Atlas of Cyberspaces" [19] provides a beautifully illustrated overview of this area. The well-known AlphaWorld Mapper (<http://mapper.activeworlds.com/aw/>) by Greg Roelofs and Pieter van der Meulen provides access to an impressive zoomable 2D map of a virtual world that is roughly the size of the state of California (429,025 km<sup>2</sup>) [20]. However, the scale of hardware and other resources required to generate such a map is beyond the reach of most virtual world researchers. Until recently, it was not possible for researchers and users of virtual worlds to create a map of a virtual world or to analyze and visualize user interaction data collected in virtual worlds. The toolkit described in this chapter allows researchers to do this on a regular basis and in a consistent manner. The results have been shown to provide new insights about the evolution and usage of virtual worlds and the activities of their users [8, 21, 22].

### **3. Data in Virtual Worlds**

It has been estimated that 80% of all data that exists in our world today has a spatial component [23]. This suggests that using spatial analysis and visualization techniques is essential in order to gain a fuller understanding of social patterns in real as well as virtual spaces. However, in the real world, collecting information about pedestrian flow, traffic or building utilization is difficult and often requires invasive and expensive methods such as the placement of sensors or cameras at various locations. The task of data collection is greatly simplified in virtual environments due to the ease with which user behavior can be recorded in a manner that does not interfere with the activities of inhabitants of the virtual world. The monitoring methods used in virtual worlds also have the inherent advantage that they result in data that is clean and consistent across time and space. The data therefore readily lends itself to analysis using data mining methods. In particular, the ActiveWorlds SDK (<http://www.activeworlds.com/sdk/>) offers ways to collect information such as movement, orientation, chat text, interactions such as clicks and even gestures using programmed software agents known as "bots". Information on the structure of virtual worlds is obtained using what are known as "prodump" and "registry" files, which respectively describe relative and absolute geometry

of objects in the virtual world. The availability of such rich data eases the study of the interplay among space and social behavior. However, before one can begin addressing these questions, one must first identify the subjects of such a study, i.e., the user groups and the tasks that they perform within or related to the environment under consideration. This is the focus of the next section.

#### 4. User Groups and Tasks

Our work in virtual worlds is embedded primarily in educational settings. Our users consist of kids from primary schools exploring information in the virtual world, performing scientific experiments, taking part in team-based games and engaging in other social activities—always with *learning* as the underlying goal. In this context, we aim to support three user groups:

*Inhabitants:* These are the core users who actually use the virtual world for activities such as exploring an art exhibition or participating in a treasure hunt. In our case, this usually consists of school kids who take part in the learning activity and their teachers who act as moderators or guides. This user group has all the needs and problems associated with using virtual worlds for collaborative work such as navigation, communication and coordination.

*Designers:* These are professionals who design the virtual world, set up the “infrastructure” and mentoring strategies to perform various kinds of activities. They also configure the environment so that it is suitable for diverse events to be held in the virtual world. This user group consists of technical designers who use computer and artistic skills to create the virtual world as well as teachers who aid the designers in creating an environment that supports a certain learning goal. This user group defines the overall design of the virtual world; placement of objects, locations of teleports, colors, textures, and behaviors of objects in response to user actions such as clicking or movement. This user group can benefit from answers to questions pertaining to patterns of usage in virtual worlds, such as the influence of the age of buildings in the virtual world on their usage, characteristics of objects that attract a high amount of interaction versus those that are ignored, etc.

*Researchers:* This user group is primarily concerned with asking and answering questions about social and technological aspects of virtual worlds such as “What are the effects of the environment on user behavior?”, “What is the extent of influence of interface devices on the navigation capabilities of users?” and “What role does spatial reference play in the interaction between avatars?”.

As is evident, each of these user groups has distinct roles to play either inside virtual worlds, outside or both. The specific nature of problems that each user group encounters varies widely. However, one can see that answering any of

these questions needs information that synthesizes data about local and global patterns of activity. For example, in order to understand how often particular areas of the world are used, one needs to step back and look at the global pattern of density of usage of the virtual world. At the same time, to understand why such a pattern exists, one needs to zoom-in to specific regions of the world and examine interaction data or chat text to identify what the avatars in those regions have been exploring or talking about. Observations about such local and global activity can then be combined with the analysis of learning outcome but also pre and post-test questionnaires to obtain a comprehensive picture of a virtual world event.

## 5. Patterns of Social Activity

The term “pattern” can be defined as: a perceptual structure, a customary way of operation, a model worthy of imitation, a decorative or artistic work. Patterns occur in nature due a variety of physical, chemical and biological mechanisms. Many of these patterns manifest themselves as concrete physical phenomena and lend themselves to immediate perception in a sensory manner. At the same time, we find a number of patterns that are hidden, sometimes due to their existence on a plane other than the physical or sensorial, requiring a serious cognitive effort to uncover them, at other times due to their recursive nature, which reveals only the gross manifestation, while the intricacies are hidden away within the structure. Curiously, patterns of social activity in virtual worlds may be found to be of both types—visible as well as obscure. Geospatial patterns such as patterns of user trails are highly visible and fairly simple to visualize, while patterns in semantic space such as those found in chat activity are more difficult to uncover.

### 5.1. Social Diffusion Patterns

In a virtual world, avatars perform activities in isolation or in the immediate presence of (or collaboration with) other avatars. We use the term *group* to denote a set of users that perform an activity together. The characteristics that define a group are similar to those necessary for the existence of a community [24], including “a common interest” and “being rooted in the same geographical space”. However, in educational settings, a group is usually a short-term congregation of avatars who share a common goal. For example, the group of all researchers exploring ways to improve a virtual world for education form a community, while the set of students that take part in a short-term science experiment, form a group within the context of the experiment. A lone user exploring a virtual world in a random manner does not share a common purpose

with other avatars and hence does not form part of a group in our sense. Note that this notion of a *group* is different from the way we refer to it in the term *user groups*.

We define social activities as activities that are carried out by a group of avatars as part of a larger objective. In our research, this larger objective is usually *learning* and the activities may be a play, a college-recruitment event, a treasure hunt, etc. The patterns that emerge as a result of these activities, hence, are also social in nature.

The Webster's New Millennium Dictionary of English defines the term *diffusion* in several ways, one of them being "the process by which a cultural trait, material object, idea, or behavior pattern is spread from one society to another". For such a diffusion process, two things are necessary: (1) the diffusing element and (2) the diffusion medium. In the above definition, the cultural trait, material object or idea is the diffusing element. This is the element that actually propagates through society possibly changing the structure of society as it diffuses. The diffusion medium in the above case is society itself, which consists of a large number of discrete units (people) that all share some common characteristic that enables the diffusion of ideas or material objects. Although homogeneity of the diffusion medium is not a pre-requisite for a diffusion process, a heterogeneous diffusion medium presents a complicated case and hence, here we consider diffusion only in a homogenous diffusion medium. We can then classify types of diffusion on the basis of whether the avatars of the virtual world act as (1) the diffusing element, or (2) the diffusion medium.

#### 5.1.1. *Users as Diffusing Element*

Users in a virtual world move in diverse ways in accordance with their personal goals or the requirements of their group. Members of a group of users move towards areas that are most suitable to their tasks. For short-term activities, this process is not evident, since the quick succession of movements and teleports can hardly be termed diffusion. However, in the case of large worlds such as the *Avatars* world (see <http://ccon.org>), which is 2 kilometers long and 2 kilometers wide, we see that over time, there is a non-uniform distribution of users through the world due to certain areas of the world being used more often than others. Such a pattern is also reflected in the evolution of virtual world buildings over time, where we see distinct building patterns that seem to suggest that certain parts of the world are more preferred than others (figure 3-1).

#### 5.1.2. *Users as Diffusion Medium*

An interesting way of looking at user behavior in virtual worlds is obtained by viewing users of virtual worlds as actors and a medium for the diffusion of ideas. Users in a virtual world communicate with each other in a variety of ways,

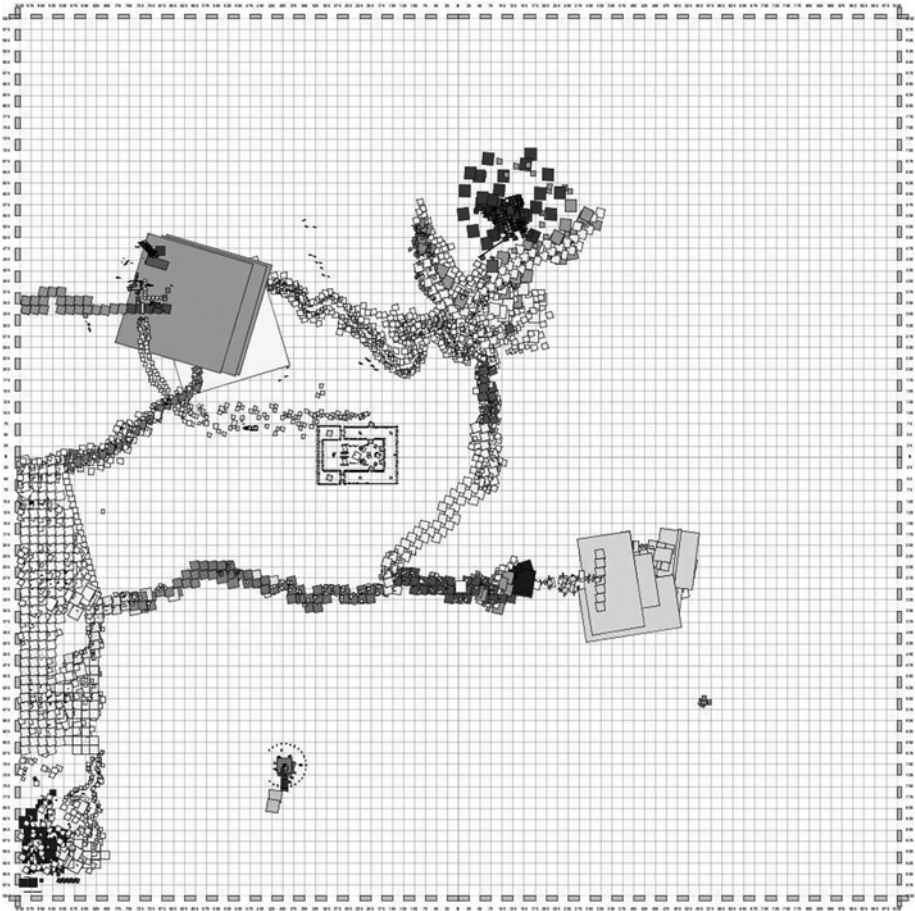


Figure 3-1. Map showing the structure of the “Avatars” world in the ActiveWorlds universe. Grid lines are used to indicate the size of the virtual world: the distance between two grid lines corresponds to 25 meters in the virtual world. The rectangles represent buildings and other structures in the virtual world and are color-coded by age. Older buildings are shown in darker color while younger buildings are shown in lighter color.

including whispering, chatting, gesturing, etc. with chat being the dominant method of communication. In fact, many users show very little movement during the entire duration of their visit to a virtual world and restrict their activity almost exclusively to chatting.

Today, virtual worlds offer a variety of ways to control the effect of the environment on the users and their activities; among the most common being configurable limits on the “visibility” of objects and the “hearing range” of chat. Visibility places a limit on how far an avatar can see in the virtual world. Objects that are farther away can be made less visible (by blurring them or hiding them by fog) or not visible at all, while objects that are closer to the



avatar are made clearly visible. Hearing range limits how far away an avatar can move from a sound source (or an avatar uttering chat phrases) before it stops “hearing” sounds or receiving chat text. An avatar must be within a particular distance from another avatar to be able to receive chat-text from the latter. These limits ensure that awareness of events in a virtual world is not *broadcast* to all the users; there is clearly a diversity of awareness or knowledge of events among users. Thus, if an idea or a concept (or simply a word) appears in the chat text of one user, it spreads to the other users subject to the limitations posed by visibility and hearing range. This spread will depend—as with any diffusion process—on a variety of factors including the density and distribution of users, the “strength” of an idea or its relevance to subsets of users. Here the idea becomes the diffusing element, while the avatars become the diffusion medium.

## 5.2. *Emergence of Patterns in Group Situations*

There is a great deal of diversity in the kinds of groups that are found in virtual worlds and they vary from short-term task based groups such as the ones we encounter in our research—to longer-term communities such as the “E-Church” world studied by Schroeder, Heather and Lee [25]. Research on groups in virtual worlds is in its infancy and much research is needed before we will achieve a comprehensive picture of groups, roles and interactions in virtual worlds. This task is not simple because groups in virtual worlds have different dynamics compared with groups in the real world. Therefore, one cannot assume that communities in virtual environments are a mere projection of communities in the real world [26]. Keeping this in mind, we caution the reader to view the subsequently discussed types of user groups as a personal observation coming out of looking at large amounts of virtual world data, rather than a complete list.

*Leader-Follower:* This type of group is characterized firstly by the presence of a single avatar that is automatically accepted as the leader and secondly by the existence of prescribed paths and actions. Members of the group (not including the leader) are the followers which obey orders, follow instructions given by the leader or simply copy her actions. Some examples of such a group include a group taking a guided tour (the tour guide being the leader) or a group of students working on a building project (the leader of the group being their teacher). The followers in the group are not expected to significantly deviate from the pattern set by the leader.

*Moderated:* This type of group is different from the one outlined above in that there are no real followers in the group. The group consists of autonomous members free to do as they wish within the scope of the activity which the group is involved in. In a way, the moderator *could* be considered the leader

of the group, since the moderator holds the right to decide upon the range and variety of actions that can fall within the scope of that particular group activity. An example of such a group is a research discussion group on virtual worlds led by the organizer of the event. In this particular group, researchers are free to discuss issues related to research in virtual worlds. However, if a heated debate gets underway or if the group begins to digress away from the main theme of discussion, the moderator can step in and restore order and focus. Typically such groups also possess the trait that its members all share a personal goal that is a subset of or is related to the common goal of the group. For example, every virtual world researcher has his/her own research agenda. However, a common thread of interest in virtual worlds binds the group together.

*Competitive:* This type of group is characterized by the presence of sub-groups within the main group, each of which competes against the other for some reward. All members of the sub-group share with each other a sense of belonging and a group identity. At the same time, each sub-group member maintains a sense of separation from the competing sub-groups, while they may all belong to the same super-group. In the geospatial sense, this group is usually seen to have fixed destinations but variable paths as each sub-group takes varying approaches to achieve their goals. A treasure hunt played by kids of a class is an example of a situation where such groups are encountered.

*Ad-hoc:* This group is characterized by its short-term existence. Members of this group do not usually possess a shared objective outside the context of the group activity being performed. For example, in the case of a group of school kids taking part in a role play, the kids may or may not have a common goal related to stage-performance. In this particular case, the group is brought together for the sake of the play and is usually dissolved as soon as the task ends. Each child in the group associates himself/herself with the group activity only as long as the activity is being carried out.

The traits described above are not mutually exclusive. Two avatars can collaborate in order to perform a task and this group might be competing against other similar groups, all of which belong to the one bigger group that is led by a single avatar. Also one might notice that the above listed traits result in progressively more decentralized groups. The level of decentralization and more importantly the concurrence of objectives of the members of the group can play a significant role in the paths taken by members of the group during an activity. Geospatially speaking, the trail of a member of a group or the group as a whole can vary depending on how coherent the behavior of the members of the group is with respect to each other and the group as a whole. Looking at the patterns of avatar trails in virtual worlds with this idea in mind reveals that we can classify group trails into the following four types in terms of the variation of their geospatial distribution over time [8] (see figure 3-2):

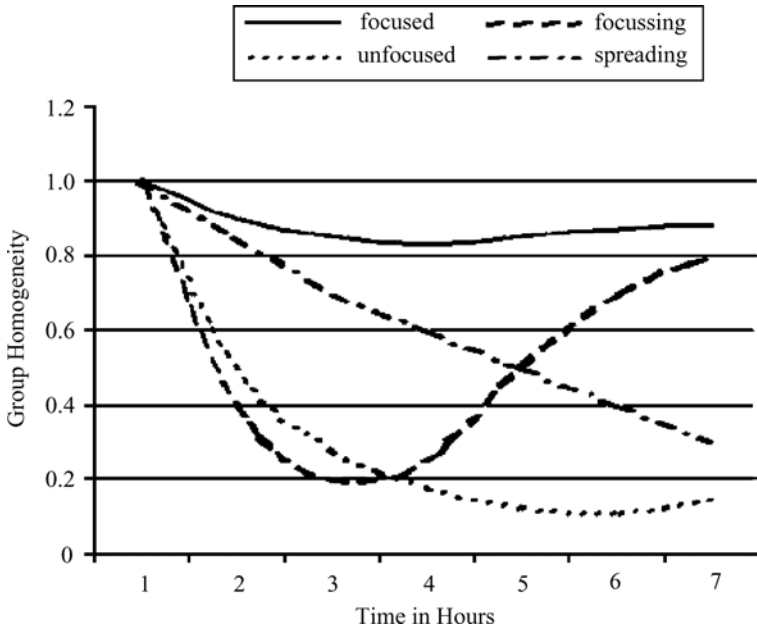


Figure 3-2. A hypothetical example showing the diffusion behavior of groups of avatars. The four types of behavior can be represented by a function that captures the variation of group homogeneity over time.

- *Focused*: This group remains tightly knit and the members of the group do not deviate much from each other's paths.
- *Unfocused*: The members of this group move as they wish, resulting in quite random trail patterns.
- *Focussing*: This group starts out dispersed all over space but over time tends to congregate as the trails of the group members become increasingly similar to each other.
- *Spreading*: This group starts out with all its members on a common trajectory, but over time, the members of the group diverge, resulting in very diverse trail patterns.

One can now see that specific kinds of events can result in specific types of diffusion behavior. An art exhibition where all visitors follow pre-set paths around the exhibits results in focused group trails. A random group of users in a public virtual world results in unfocused diffusion behavior, as each member explores the world individually. An event where members of the group are assigned set regions of the world which they scan results in focusing or spreading group trails. Identifying the type of group trail resulting from an event can give researchers a good idea of whether the event went as planned. This idea is elaborated in the next section.

## 6. Analysis and Visualization

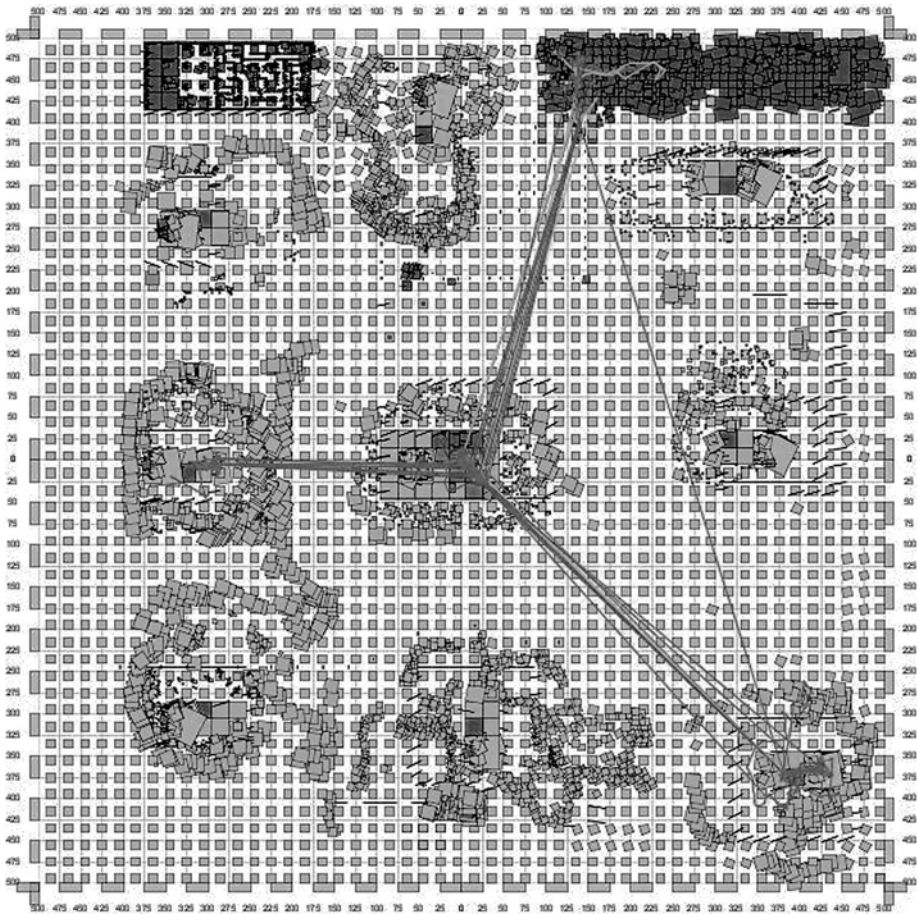
Any study on patterns requires tools for repeated and consistent analysis of data. A good toolkit can facilitate this by providing researchers with a wide variety of perspectives from which to analyze and view a data set. An effective visualization can summarize the data, highlight anomalies or outstanding features and even reveal patterns that are impossible to identify by looking at data in its raw form. In this section we describe how visualization can support the study of social behavior.

### 6.1. *The ActiveWorld Toolkit*

The ActiveWorld Toolkit (available for download at <http://ella.slis.indiana.edu/~sprao/research/virtualworlds/>) is a free toolkit that helps researchers to analyze and visualize world structure and user behavior data in virtual worlds built with ActiveWorlds technology. The toolkit allows one to load world data (*propdump* and *registry* files) and data about activity of avatars in the virtual world (bot log files). The world data is usually obtained by using an administrative utility for the virtual world which can export data about location, orientation, size and information on how the objects react in response to user actions such as movements or mouse clicks. User data is collected in the virtual world using bots which continuously monitor user activity such as movements, chats, clicks, teleports, etc.

To analyze how this toolkit can help researchers analyze virtual world events, consider an educational treasure hunt in a virtual world where the idea is to let kids follow clues to find treasures, picking up bits of information about issues such as environmental pollution, conservation of natural resources, etc. This requires designing the clues for the hunt in a way that is appropriate to the level of the users' cognitive abilities. One cannot incorporate clues so complex that kids get frustrated and quickly lose interest. On the other hand, the clues must be complicated enough to ensure that not all kids immediately figure them out. Hence, the ideal set of clues for the hunt would help the kids "stay in the flow" [27] of the hunt.

Also critical to the success of such an event is the ability to successfully navigate through the world. Navigational clues must be placed in strategic locations to ensure that the kids can find their way around the world, but at the same time they should not give away the locations of the clues themselves. Some questions about the treasure hunt that can be asked are "Were most of the users able to follow the clues and find their way through the world?", "Which group of users was most focused?", "Which users were lost and at what point?", "Which parts of the treasure hunt were most or least challenging?" etc. These are some questions which can be answered using visualization of user data.



*Figure 3-3.* Visualization of the trails of the participants of three parallel Spanish learning sessions in LinkWorld. The three sessions are taking place in the north, south-east and west. The trails of the participants (avatars) are represented by lines. The rectangles represent buildings and other structures in the virtual world.

An example of such a visualization is shown in figure 3-3. This visualization shows the trails of users that took part in three parallel Spanish learning sessions in LinkWorld, with the treasure hunt being held in the northern part of the world. This map gives us an overview of where avatars went and where they have been most active, i.e., where they have chatted the most or where they have clicked and interacted with objects the most. In order to answer specific questions, however, one needs to be able to examine the interactions of each individual user. This is facilitated in the toolkit using the zoom function, which allows one to focus on the area of interest. The toolkit allows one to smoothly move between overview and detail mode, thus helping the user maintain her orientation while working with the toolkit.

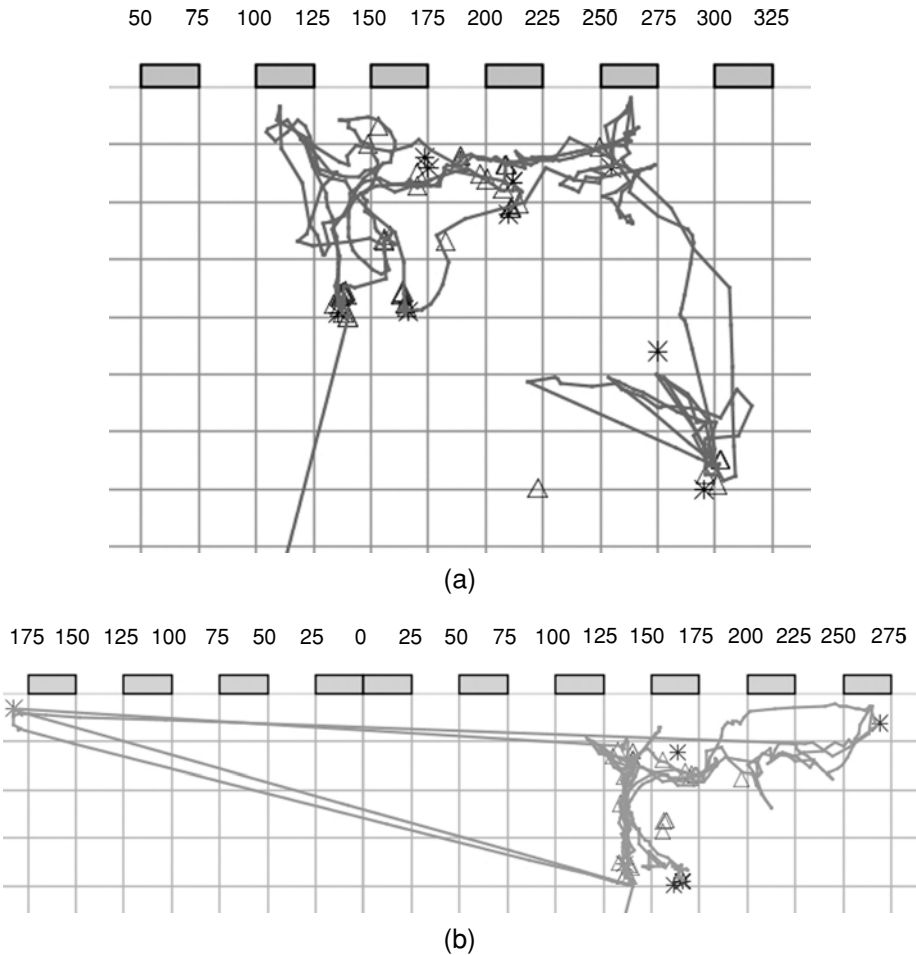


Figure 3-4. Two participants who have strayed away from the main treasure hunt area during the event. Triangles show locations where participants chatted, stars show locations where they clicked objects in the world, while trails show the path that these two participants took. The numbers at the periphery indicate the distance in meters from the centre of the virtual world along the x-axis.

## 6.2. Analysis of Geo-Spatial Patterns

Figure 3-4a and b show zoomed-in views of the trails and chats (triangles) of two users who seem to have strayed significantly from the area where the treasure hunt was being held. In both cases an examination of the chat text revealed that the moderator of the treasure hunt realized that two participants were lost and made an attempt to bring them back. One can also look at the chat text of the participants to understand when they themselves realized that they were not in the right area. There are several reasons why such confusion could have arisen: the clues for the hunt were not right, or perhaps the placement of

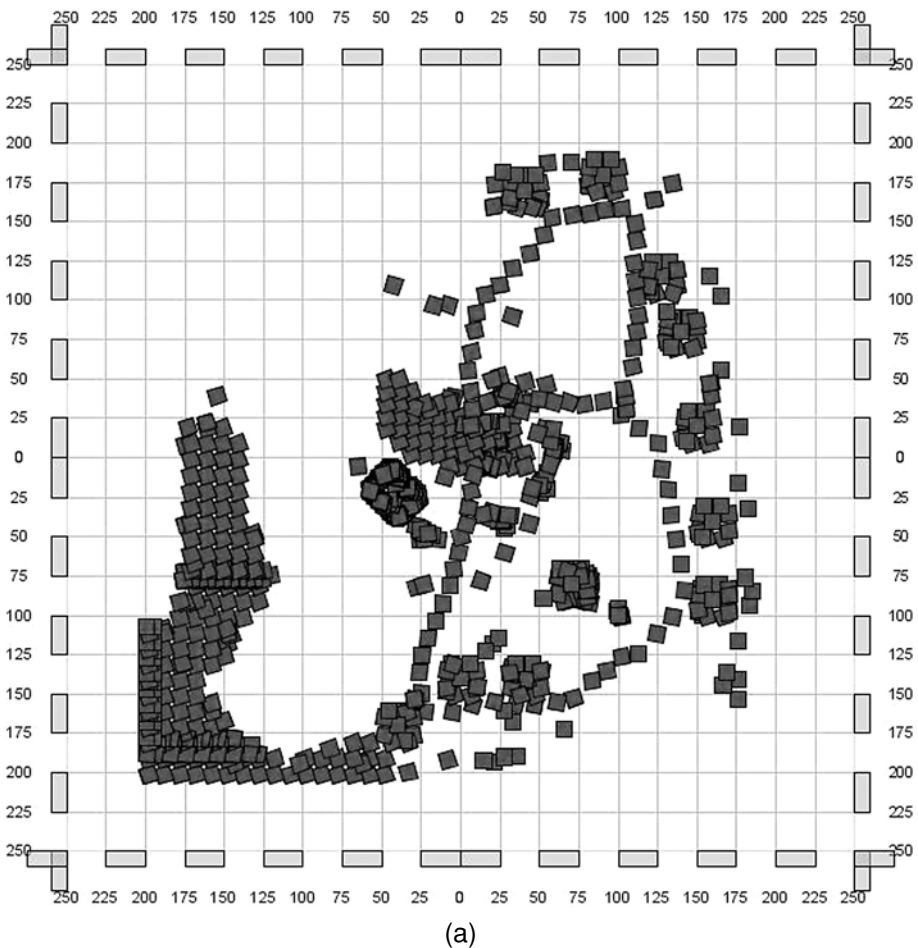


Figure 3-5. Three views of an art show in the Avatars world. (a) Structure of the world showing positions of objects. (b) Trails of avatars viewing the art show (represented by lines). (c) Locations where users chatted (triangles) and clicked objects (stars).

objects was such that the participants were misled into following a path that was out of bounds for the event, or maybe the participants were curious to explore other areas in the world. A conclusion of the latter type is especially useful when a statistical analysis is planned; one can quickly decide whether or not the behavior of the user in question is an anomaly and accordingly exclude or include him/her in the analysis. A quick peek at the chat text of these avatars revealed that one of them (figure 3-4a) spent a considerable amount of time in an area out of bounds of the actual hunt before coming back to the main area, while the other avatar (figure 3-4b) was brought back to the main area almost immediately after they left the group.

Figure 3-5 shows another set of visualizations that uncover behavioral patterns in virtual worlds. These are visualizations of an art show held during the

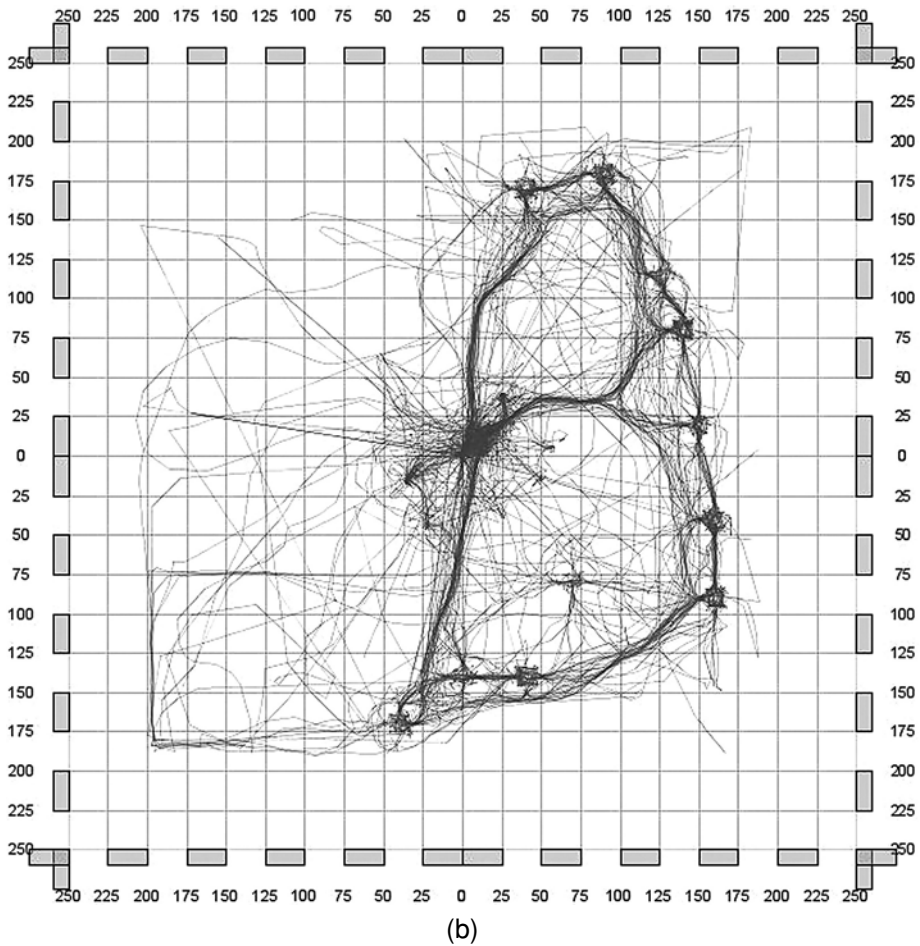


Figure 3-5. (Continued)

Avatars! Conference in 2002 (see <http://www.ccon.org>). Figure 3-5a shows the structure of the world. The art exhibits are placed roughly in the shape of the letter “B”. Figure 3-5b shows the trails of the avatars as they explored the exhibition. From these two figures, it is clear that the avatars followed the path of exhibits very well. The trails which deviate from the center were found to belong to avatars who, after seeing all the exhibits, decided to explore the world a bit more and then exit. Finally, figure 3-5c shows the locations of chat (triangles) and clicks (stars) overlaid over the structure of the world (dark objects). This final figure gives us a clear picture of the exhibits near which people chatted with each other or interacted with objects. This information can be used along with post-event questionnaires to determine what aspects of those locations or exhibits encouraged people to talk to each other.





(c)

*Figure 3-5. (Continued)*

### 6.3. *Analysis of Chat Data*

Chat data in virtual worlds is especially interesting because it can potentially reveal patterns about users' thought processes while they performed activities inside the virtual world. Chat data collected over a long period of time give us a picture of the evolution of the topics discussed among the inhabitants of the virtual world. A particularly interesting analysis is the detection of changes in the frequency of usage of particular words. The results enable us to make inferences about the emergence of topics in a conversation among avatars.

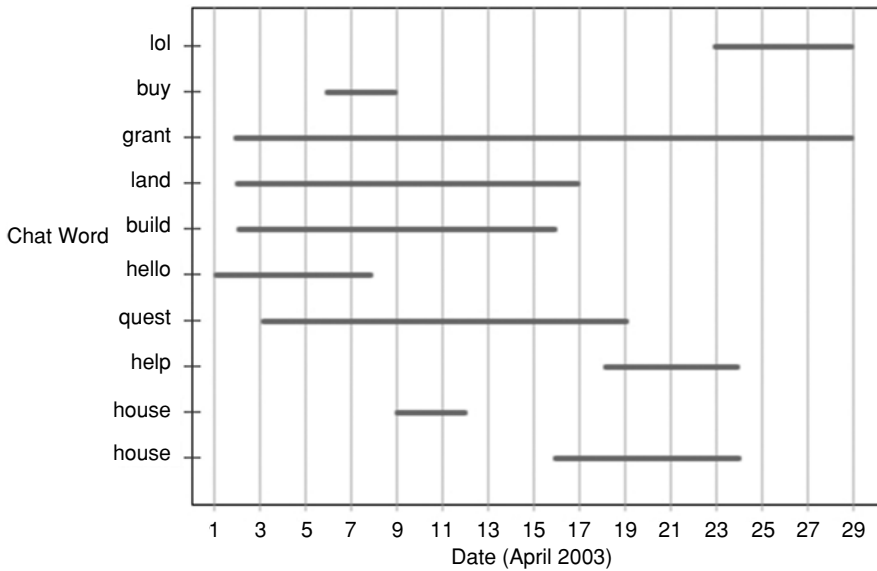


Figure 3-6. Top ten bursty words over the month of April 2003 in the Culture world of the Quest Atlantis universe. The vertical ( $Y$ ) axis shows the burst words, while the horizontal ( $X$ ) axis shows the dates during which the bursts occurred. For each word, a horizontal line is drawn starting from the date that the burst began till the date that burst ended.

Such an analysis demands that we do not restrict our search to any particular set of words, but rather that we let the analysis itself determine the set of most popular words. Specifically, we are looking for words which suddenly “burst on the scene”, generate a lot of activity and then die down. We believe such *bursty* words are good indicators of popular topics in chat.

In order to identify bursty words, we can use a predictive model such as Kleinberg’s burst detection algorithm [7]. Simply speaking, this algorithm considers a word to be bursty if it appears with a high frequency over a time period that is short compared to the total time span of the complete data set. Note that the set of words identified using this analysis are not simply the set of words which have high frequency of usage; in fact a word that constantly keeps appearing in chat will not be recognized as a bursty word.

Figure 3-6 shows the result of such an analysis performed on chat data collected over the month of April 2003 for the Culture world in Quest Atlantis universe (<http://questatlantis.uni.activeworlds.com>). The “ $X$ ” axis shows the bursty words. The “ $Y$ ” axis represents the dates in the month of April. Each horizontal line begins at the date at which the word started to burst and ends at the date at which the usage of that word died down. Table 3-1 shows the strength of the burst for these words. The “strength of burst” indicates the amplitude of the burst of that word, i.e., how intense the usage of that word was. Table 3-1 shows us that the word “lol” (an acronym for Laugh Out Loud), appears with very high

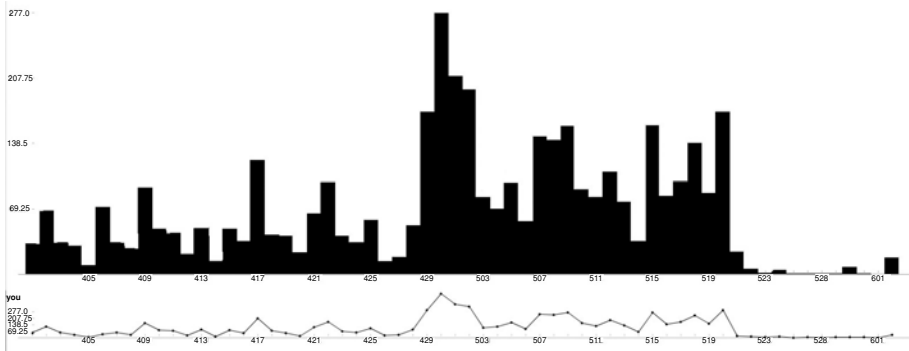
Table 3-1. Strength of burst of the top ten bursty words in chat data over the month of April 2003 in Culture world in the Quest Atlantis universe.

Word	lol	Buy	Grant	Land	Build	Hello	Quest	Help	House
Strength of burst	88.213	50.387	45.257	33.25	31.555	30.492	29.879	29.481	28.694

amplitude compared to other words such as “grant” or “hello”. However, the latter category of words have a longer life-span: the word “lol” remains in the conversation for 7 days but the word “land” and “build” remain dominant for 15 and 16 days, respectively. Note that the word “house” appears twice, meaning that this word appeared as the dominant word twice in one month—the first time between April 9 and 12 and the second time between April 16 and 24.

From figure 3-6, we know exactly what day the burst of a particular word occurred and therefore we can go back to the original data and determine the cause for that burst. In this case, the reason for the sudden burst of the word “lol” was traced back to a kid who kept welcoming every new visitor into the virtual world with strange and funny welcome messages. The messages were written in a way that made them look as if they were coming from an automated bot. Every time this kid made a fool of a new user, everyone assembled in the virtual world would burst out in laughter. The other words have a more serious tone and clearly indicate building and trading activities going on very actively for a period of about 15 days. One can also take advantage of the starting and ending dates of the burst and make inferences about the chain of events that took place. So we see that “hello” is one of the first words that burst—probably a time for introductions—after which users got down to business talking about “quest”, “build” and “land”, ending with “lol”. One can see that simple analyses like these can lead to interesting inferences that can then be verified using questionnaires and interviews.

The sequence of chat data can be visualized to show trends in the rise and fall of the word frequency over time. One of the most frequently used words in chat text which did not appear in the analysis outlined above was the word “you”. This word was consistently used with very high frequency and hence was ignored by the burst algorithm. Figure 3-7 shows the frequency of usage of the word “you” over time in the same data set analyzed above visualized using TimeSearcher [28]. From the figure, the pattern of usage over time is immediately evident: the word started off at a low frequency, its average increased slowly over time, it took a big jump after the half-way point through the time period of observation and maintained a relatively high average frequency until finally the frequency dropped very suddenly. This pattern of frequency can then be compared to changes in the user group, environment or activity in the virtual world to yield additional clues.



*Figure 3-7.* Visualization of the frequency of usage of the word “you” in a two-month sample of chat data in Culture world in the Quest Atlantis universe. The vertical axis ( $Y$ ) indicates the frequency of usage of this word, while the horizontal ( $X$ ) axis indicates time. The bar graph (top) and the line graph (bottom) provide the same information. However, in the TimeSearcher application, the bar graph allows users to zoom in and out as well as query for details over a particular time period, while the line graph acts as a constant reference.

## 7. Outlook

In this chapter, we have seen how one can combine diverse methods of analysis and visualization in order to create powerful ways to analyze user behavior data and discover patterns in them. The use of patterns for social research in virtual worlds is not only powerful, but necessary in order to truly understand the impact of space and time on the behavior of individual users and groups in virtual worlds. Visualizations provide researchers with effective means to discover these patterns and enable efficient communication of analysis results to peers.

In the future, we hope to apply these analysis and visualization techniques to gain an understanding of virtual world behavior in real-time. One can imagine a tool that provides continuous real-time summaries of activity patterns of users as they perform tasks in a virtual world. Such a real-time map could provide navigational support to users, present a real-time summary of chat topics emerging in different areas of the virtual world, or even display a dynamically-generated social network of avatars as they interact in the virtual world. We believe that such tools will prove to be essential for determining the influence of spatial, semantic, and social factors on dynamic group behavior.

Another line of research that looks promising is the development of techniques to automatically identify roles that avatars take in a group based on spatial patterns, chat utterances and interaction data. For example, we can ask the question, “From the patterns of social behavior, is it possible to identify automatically which avatar emerged as the leader?” Such a question calls for an

intelligent synthesis of theories and techniques developed in computer supported cooperative work, information visualization, spatial analysis, clustering methods, computer-mediated communication [29], social network analysis [30], and others.

### Acknowledgements

We would like to thank Bonnie DeVarco of the BorderLink project (<http://www.borderlink.org>) for her support throughout this research and for access to the LinkWorld data. Thanks also to Sasha Barab and Bob Cartheaux of the Quest Atlantis project (<http://questatlantis.uni.activeworlds.com>) for their valuable comments and for permitting us to log Quest Atlantis log data and Bruce Damer of Digital Space Commons (<http://www.digitalspace.com>) for granting us permission to log the “Avatars!” Conference and generate maps of the Avatars world. Thanks also to Rich Bernstein of Cornell University whose helpful suggestions led to several improvements to the ActiveWorld Toolkit. This work is supported by a National Science Foundation grant under Role-0411846.

Figure 3-2 reprinted with permission from “Social Diffusion Patterns in 3D Virtual Worlds”, *Information Visualization Journal*, Vol. 2, Issue 3, Sept. 2003. Copyright Palgrave-Macmillan, UK.

### References

1. Chen, C. (1999). *Information Visualisation and Virtual Environments*. London: Springer Verlag.
2. Card, S., Mackinlay, J., & Shneiderman, B. (Eds.) (1999). *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufmann.
3. Ware, C. (2000). *Information Visualization: Perception for Design*. Morgan Kaufmann Interactive Technologies Series. San Francisco: Morgan Kaufmann.
4. Spence, B. (2000). *Information Visualization*. Reading, MA: Addison-Wesley.
5. Börner, K. (2002). Twin worlds: Augmenting, evaluating, and studying three-dimensional digital cities and their evolving communities. In M. Tanabe, P. van den Besselaar, & T. Ishida (Eds.), *Digital Cities II: Computational and Sociological Approaches*. Japan: Springer Verlag, pp. 256–269.
6. Rabiner, L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, 77(2), February, pp. 257–286.
7. Kleinberg, J. (2002). Bursty and hierarchical structure in streams. In *The 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, July 23–26.
8. Börner, K. & Penumarthy, S. (2003). Social diffusion patterns in three-dimensional virtual worlds. *Information Visualization*, 2(3): 182–198.
9. Aref, H., Charles, R.D., & Elvins, T.T. (1994). Scientific visualization of fluid flow. In C.A. Pickover & S.K. Tewksbury (Eds.), *Frontiers of Scientific Visualization*. Hoboken, NJ: Wiley Interscience.

10. Senger, S. (1996). Digital cadavers (TM): An environment for the study and visualization of anatomic data. In *Proceedings of the Visible Human Conference*, National Institutes of Health, Bethesda, Maryland, p. 127.
11. Dorigo, G. & Tobler, W. (1983). Push pull migration laws. *Annals Association of American Geographers*, 73(1): 1–17.
12. Plaisant, C., Milash, B., Rose, A., Widoff, S., & Shneiderman, B. (1996). *Life lines: Visualizing personal histories*. In the *ACM CHI '96 Conference Proceedings*. New York: ACM Press, pp. 221–227.
13. Donath, J.S., Karahalios, K. & Viegas, F. (1999). Visualizing conversation. *Journal of Computer Mediated Communication*, 4(4). Available at <http://www.ascusc.org/jcmc/vol4/issue4/donath.html>
14. Hochheiser, H. & Shneiderman, B. (2001). Using interactive visualizations of WWW log data to characterize access patterns and inform site design. *American Society for Information Science*, 52(4): 331–343.
15. Donath, J.S. (1995). Visual Who: Animating the affinities and activities of an electronic community. In *Proceedings of the Third ACM International Conference on Multimedia*, San Francisco, CA, pp. 99–107.
16. Smith, M., Farnham, S., & Drucker, S. The social life of small graphical chat spaces. In R. Schroeder (Ed.), *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*. London: Springer, pp. 205–220.
17. Naper, I. (2000). System features of an inhabited 3-D virtual environment supporting multimodality in communication. In *Proceedings of the 34th Annual Hawaii International Conference on Systems Sciences*. Maui, Hawaii: IEEE Computer Society.
18. Whyte, W.H. (1980). *The Social Life of Small Urban Spaces*. Washington, DC: The Conservation Foundation.
19. Dodge, M. & Kitchin, R. (2001). *An Atlas of Cyberspaces*. Reading, MA: Addison Wesley.
20. Schroeder, R., Smith, A., & Huxor, A. (2001). Activeworlds: Geography and social interaction in virtual reality. *Futures*, 33: 569–587.
21. Börner, K., Penumarthy, S., DeVarco, B. J., & Kerney, C. (2004). Visualizing social patterns in virtual environments on a local and global scale. In *Digital Cities 3: Local information and communication infrastructures: Experiences and challenges*. To be published by Springer Verlag.
22. Börner, K., Jun Lee, G., Penumarthy, S., & Jones, R. J. (2004). Visualizing the VLearn3D 2002 conference in space and time. In *Visualization and Data Analysis*, San Jose, CA. SPIE-IS&T, Vol. 5295: 24–32.
23. *Federal Geographic Data Committee, Geospatial Standards*. Available at <http://www.fgdc.gov/publications/documents/standards/standards.html>
24. Becker, B. & Mark, G. (2002). Social conventions in computer-mediated communication: A comparison of three online virtual environments. In R. Schroeder (Ed.), *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*. London: Springer, pp. 19–62.
25. Schroeder, R., Heather, N., & Lee, R. (1999). The sacred and the virtual: Religion in multi-user virtual reality. *Journal of Computer-Mediated Communication*, 4(2). Available at <http://www.ascusc.org/jcmc/vol4/issue2/schroeder.html>
26. Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M., & Haythornthwaite, C. (1996). Computer networks as social networks: Collaborative work, telework, and virtual community. *Annual Review of Sociology*, 22: 213–238.
27. Bederson, B. (2004). Interfaces for staying in the flow. *Ubiquity*, (5)27, Sept. 1–7. Available at <http://www.acm.org/ubiquity/views/v5i27.bederson.html>

28. Keogh, E., Hochheiser, H., & Shneiderman, B. (2002). An augmented visual query mechanism for finding patterns in time series data. In *Proceedings of the Fifth International Conference on Flexible Query Answering Systems*, October 27–29, Copenhagen, Denmark. *Lecture Notes in Artificial Intelligence*. London: Springer.
29. Herring, S.C. (2002). Computer-mediated communication on the Internet. *Annual Review of Information Science and Technology*, 36: 109–168.
30. Wasserman, S. & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.