**Identifying and ranking stability, topological significance, and redundancies in water resource networks to guide detailed simulation modeling**

Leah Meeks, S.M. ASCE, M.E., EIT1 and David E. Rosenberg, M. ASCE Ph.D2

1 PhD Candidate, Dept. of Civil & Env. Engineering and Utah Water Research Laboratory, Utah State University Dept. of Civil & Env. Engineering, 4110 Old Main Hill, Utah State University, Logan, UT 84322-4110; c:(435)-881-8697, [l.meeks@aggiemail.usu.edu](mailto:l.meeks@aggiemail.usu.edu);

2 Associate Professor Dept. of Civil & Env. Engineering and Utah Water Research Laboratory, Utah State University Dept. of Civil & Env. Engineering, 4110 Old Main Hill, Utah State University, Logan, UT 84322-4110; o:(435)-797-8689, f:(435)-797-1185, [david.rosenberg@usu.edu](mailto:david.rosenberg@usu.edu).

**Abstract**

The size and complexity of water resources networks typically require a large number of computationally intensive simulations to test effects of changes in network structure or management. Building on the qualitative definitions of node performance in a network and the less-computationally intensive parallel coordinate plot visualization (NEVIS) method introduced by Singer and Greenshpan (2009), this paper introduces a Ranking Automation for NetworKs (RANK) tool that automates the process to quantify and rank nodes that are (1) stable: their roles do not depend on the existence of particular nodes, (2) topologically significant: when removed or added to the network, these nodes cause other nodes to be unstable, and (3) redundant: node pairs that have similar connections. RANK calculates node centrality and creates a parallel coordinate plot as in NEVIS and then ranks the nodes based on the calculated pairwise differences between the elements on each plot trace. RANK is applied to the 55-node, 73-link lower Bear River water system that stretches from southern Idaho to the Great Salt Lake, Utah. Results show the least stable nodes connect to only one other node and are furthest downstream. The three most topologically significant nodes are junctions that connect large branches to the rest of the network. The most redundant node pairs have few but identical connections. The rankings identify highly redundant, low topological significant nodes as candidate sources for water transfers to urban areas (e.g., Cache and South Cache Valley agriculture). Further, investigate (a) building or expanding dams at reservoir sites with high topological significance (Mill Creek and Hyrum), (b) removing dams at sites with low topological significance, (c) adopting conservation measures or developing alternative supplies at unstable service areas (Idaho and Cache Valley Urban), or (d) monitoring flows and protecting environmental features at unstable and topologically significant nodes (e.g., various nodes along the Bear River). Future work should expand RANK to consider flow magnitude in the calculation of node centrality. The RANK tool identifies promising nodes to subsequently focus computationally-intensive simulation and sensitivity analysis efforts.

**CE Database subject headings:** Parallel coordinates, Water management, Bear River, Network analysis

**Introduction**

Water resources networks are often large and connect numerous water supply, reservoir, diversion, and demand site nodes through multiple natural and engineered conveyance links. Classical model schematics use arrows to visually show network connectivity in one figure but have difficulty conveying how each node in the network affects the other nodes as well as what network changes are potentially significant versus inconsequential. Thus, water resources analysts often make multiple computationally intensive simulation runs to test the effects of changes in network structure and management (Loucks et al., 2005). Analysts must also specify the scenarios; for large models, the computational and time demands limit the number of changes and scenarios considered. Analysts could benefit from faster network screening tools that quantify node interactions and guide the development of scenarios for more detailed study with a simulation model.

Network analysis software such as UCINET (Borgatti et al., 2002) and Cytoscape (Shannon et al., 2003) use the network schematic to quantify node characteristics such as the path between nodes, nearness (normalization of distance between two nodes), reachability (shortest path between two nodes), and density (number of one-distance links divided by the total number of possible links). These measures describe individual node attributes but do not show how nodes influence other nodes.

Singer and Greenshpan (2009) introduce a method called node extraction and visualization (NEVIS) that uses node extraction and parallel coordinate plotting to qualitatively show how nodes in networks with bidirectional links influence each other. The NEVIS method, included in Inselberg’s compilation textbook on parallel coordinates (Inselberg, 2009), systematically removes one node from a network, calculates how each node removal influences the centrality of remaining nodes (centrality is a measure of how connected a node is to the network), and plots all the extracted centrality results as ordinates on a parallel coordinate plot where each parallel axis represents an extracted node. Traces, one for each node, show how a node is affected by extracting the other nodes. Singer and Greenshpan (2009) further define the terms *vulnerable*, *topologically significant*, and *backed up* to indicate nodes that are (i) affected by removal of other nodes, (ii) affects the centrality of other nodes in the network when removed, and (iii) allow alternate paths to bypass a removed node. They also qualitatively describe how to visually identify nodes with these characteristics from the shapes of the traces on the parallel coordinate plot. Unfortunately, the qualitative descriptions do not allow an analyst to identify or compare the relative importance of nodes nor consider a large number of nodes.

Parallel coordinate plots are an attractive visualization tool because they allow the viewer to see a very large number of dimensions side-by-side in one figure and can also show relationships among variables on adjacent axes (Inselberg, 1985; Wegman, 1990). The plots contrast with traditional Cartesian coordinate plots that arrange axes orthogonally and can only show two or three dimensions. For example, a point in three-dimensional orthogonal coordinates is represented as a vector in parallel coordinates whose elements are connected by a trace of polygonal lines (Figure 1).

Elsewhere, Esdall (2003) utilized parallel coordinates to relate hurricane categorization and financial loss to multiple factors (e.g., location, time, temperature, sea level, and socioeconomic variables). The plots also helped identify errors in raw data through visual inspection for lines that did not follow the trends. Choi et al (2008) used parallel coordinate plots to visualize variables in real-time and detect anomalies that increase the likelihood of unknown large-scale internet source attacks. Albazzaz et al. (2005) compared parallel coordinate and multivariate statistical analysis methods to identify abnormal wastewater treatment operations. They tracked 38 variables for 527 days, found the parallel coordinates method required users to continuously interaction with the plot, and recommended combining multidimensional visualization and statistical method techniques to expand the science of data analysis.

These applications also found several limitations to the parallel coordinate visualization method: (1) the plots often have many lines and become busy and crowded (Edsall, 2003), (2) the ordering of axes affects the interpretation of results (Edsall, 2003, Huh and Park, 2008, Albazzaz et al., 2005), (3) they take time to construct (Albazzaz et al., 2005), (4) data analysis has not been automated (Albazzaz et al., 2005), (5) cannot compare variables on distant axes (Edsall, 2003), and (6) plots only allow for qualitative data comparisons by visual inspection.

To improve the use of parallel coordinate plots to analyze water resources networks, this paper introduces the Ranking Automation for NetworKs (RANK) tool to quantifiably define and rank nodes for three influence metrics: (1) stable (their connectivity in the network does not depend on other nodes), (2) topologically significant (they cause other nodes to be unstable when removed from the network), and (3) redundant with other nodes. RANK follows the NEVIS method to produce the parallel coordinate plot that shows extracted centrality values. However, RANK goes further to use the extracted centrality values to quantify and rank each node for the three influence metrics. The quantification and ranking control for the order extracted nodes are listed as axes on the plot, allow for unidirectional flows (e.g., water flowing downstream in a water resources network), and identify the most promising nodes on which to focus subsequent computationally-intensive simulation and sensitivity analysis efforts. The approach is used to identify promising locations to develop scenarios for agriculture-to-urban water transfers and other water management actions in a in a 55-node, 73 link network for the lower Bear River basin, Utah. Described below are the steps of RANK and its application to the lower Bear River system.

**Methods**

RANK starts by using Singer and Greenshpan’s (2009) methods to define extracted networks, calculate node connectivity and centrality in each extracted network, and build the resulting parallel coordinate plot that shows node centrality across the extracted network. RANK next defines quantitative performance metrics for Singer and Greenhpan’s (2009) qualitative definitions of stability, topological significance, and redundancy. The tool then uses the quantitative definitions to automatically rank and identify nodes of management interest. This section defines the terms, concepts, and performance metrics on which RANK is built and describes how RANK works (Table 1).

RANK has four sequential steps: 1) create adjacency matrix, 2) calculate the extracted centrality and create the parallel coordinate plot, 3) calculate pairwise differences, and (4) rank node stability, topological significance, and redundancy according to each performance metric. Each step is described and presented below.

**Step 1: Create Adjacency Matrix.** RANK uses a square adjacency matrix of rank *n* (number of nodes) to define the network topology. In this matrix, entries of one specify unidirectional links that connect a start node (row) with an end node (column). For example, if water flows from a reservoir to a city, the link points from the reservoir (start node, row) to the city (end node, column). In a network, a start node can have multiple links to different end nodes, an end node can be the start of subsequent links, and a pair of links can specify bi-directional travel between two nodes. To define the adjacency matrix, the user can either a) manually enter node names and links in a RANK worksheet or b) draw the directed graph in a program such as HydroPlatform (Harou et al., 2010) or UCINET (Borgatti et al., 2002), and then export the calculated adjacency matrix to RANK.

**Step 2: Calculate Extracted Centrality and Create the Parallel Coordinate Plot.** Next,RANK transposes the adjacency matrix prepared in Step 1 because the Singer and Greenhpan (2009) node measures are for bi-directional link relationships whereas links in water resources applications have direction and the focus is on the downstream (end) node. Then, RANK follows the NEVIS steps: a) create extracted networks by removing each node one at a time, b) calculate an extracted connectivity matrix for each extracted network, c) calculate node centralities in each extracted network, and d) create the parallel coordinate plot with extracted centrality values as ordinates, extracted nodes as abscissa, and traces that show how each node’s extracted centrality values change across the abscissa (see details in Singer and Greenshpan, 2009).

The parallel coordinate plot allows the user to visually inspect the nodes for stability, topological significance, and redundancy. Stable nodes can be visually identified as nodes whose traces have few or no vertical drops across the abscissa (extracted nodes). Topologically significant nodes are locations on the abscissa where extracting a node causes multiple traces to drop and/or large drops in traces. Nodes with similar horizontal traces are likely to be redundant. Note that the order of the axes influences this qualitative visual interpretation results.

**Step 3: Calculate Pairwise Differences.** To quantify the performance metrics, RANK next calculates the difference between each of the (n-1)(n-2) ~ O(n2) pairs of extracted network centrality values along a trace (extracted centrality is undefined on a node’s trace at the abscissa where the node is extracted). These pairwise differences permit simultaneous comparison across each parallel coordinate axis. RANK quantifies node stability as the average of all pairwise differences associated with the trace. Higher average pairwise differences indicate nodes whose traces have larger drops and are more affected by node extractions. In contrast, each node’s topological significance is determined by averaging the (n-1)(n-2) ~ O(n2) paired differences in extracted centrality values generated from the n-1 locations where traces for each other node cross the abscissa for the extracted node.

**Step 4: Rank Nodes.** RANK uses averages of the pairwise differences to determine the node ranks for stability, topological significance, and redundancy. The most stable node has the lowest average pairwise difference across a trace and describes traces that do not have large or multiple drops. Topological significance is quantified by examining two factors associated with the traces of extracted centrality values: the number of drops at an extracted network axis and the magnitude of each drop. Multiple traces that drop at the same extracted node indicate that extracting that node causes many nodes to become unstable. Extracted nodes that cause large numbers of traces to drop and large magnitude drops are topologically significant. The topological significance magnitude for a node is the average of all pairwise differences along the axis. RANK counts the number of pairwise differences for each axis which are equal to or greater than a user-specified threshold. Each node is ranked for number of and magnitude of drops. The node having the highest average rank for each is the most topologically significant.

To determine redundancy, the program identifies candidate redundant node pairs to avoid needlessly considering all possible node pairs and reduce the calculation time. A histogram is created for each node with the number of pairwise differences in each of 15 histogram intervals. Candidate redundant node pairs are classified as nodes that have less than 0.5% different for each histogram interval. RANK then compares the row vectors from the connectivity matrix for each node in the candidate pair. Redundancy is expressed as a percentage and calculated as the number of common connectivity values divided by n-1 (maximum possible number of connections).

**Implementation.** RANK uses Excel’s Visual Basic for Applications (VBA) macro programming capabilities to automate the entire analysis. Automation requires the user to provide three inputs: directed graph of the network for adjacency matrix or manual input of the adjacency matrix, a value for the parallel coordinate drop threshold to determine topological significance, and the redundancy threshold RANK uses to screen redundant pairs to show to the user. RANK produces the parallel coordinate plot, ranks each node’s stability and topological significance, and lists node pairs that are redundant. RANK can be accessed at https://github.com/lmeeks/RANK.

**Applications**

RANK is first demonstrated for two small hypothetical networks. Then, it is applied to inform management of the much larger lower Bear River water system that has 55 nodes and 73 links in Idaho and Utah.

**Illustrative Networks.** Illustrative (i) single branch and (ii) hub and spoke networks are presented because they are simple and uniform in construction but have very different structure. These networks introduce how RANK works, verify that outputs are correct, and illustrate the performance indicators that quantify stability, topological significance, and redundancy. Each network and the key results are presented below.

*Single Branch*. The single branch network has a single source node (A), multiple intermediary nodes connected by single links (B through I), and a single sink node (J) (Figure 2-A1). The adjacency matrix for the single branch network has entries of 1 just above the primary diagonal. Running RANK yields a parallel coordinate plot where each horizontal trace is different but all follow a similar trend where the centrality of each node decreases as closer, upstream nodes are removed. The most stable nodes are the most upstream in the network: A, B, and C. Because of the number and sharpness of drops in horizontal traces at nodes E, D, and F, these middle nodes are the most topologically significant. There are no redundant node pairs because each of the nodes have different connections. The quantitative results confirm what may be apparent from visual inspection that upstream nodes are the most stable while the most topological significant nodes are the nodes near the middle of the network that when removed most affect a large number of nodes immediately downstream.

*Hub and Spoke.* In a hub and spoke network, a single hub node (A) is the sole source for all exterior spoke nodes (B to J in Figure 2-B1). The adjacency matrix for this network has entries of 1 in the row for the hub node. The parallel coordinate plot shows that the hub (A) is stable as its trace has a constant value of zero, the centrality values for all spoke traces drop at the axis A which represents extraction of the hub node. Since the traces for the spoke nodes drop, the spokes are unstable. Drops occur when the hub node is extracted and make the hub topologically significant. The traces for all the spokes are identical and make the spoke nodes redundant with each other. The results confirm what is likely apparent from visual inspection that removing the hub node from a hub and spoke network decreases the stability of the other nodes the network connectivity.

These illustrative networks provide a way to verify the accuracy of RANK to both visualize and quantify stable, topologically significant, and redundant nodes in simple networks. Further, the parallel coordinate plot for the Single Branch network does not readily convey that the middle nodes are most topologically significant (extracting these nodes creates the largest and most number of drops in other traces); this example additionally shows the importance of RANK’s quantification. The results also show that the performance metrics of stability, topological significance, and redundancy are not mutually exclusive – for example nodes can be both unstable and redundant (spokes in a Hub and Spoke network) or topologically significant but not stable (mid-range nodes in Single Branch network). Each of these quantified and ranked characteristics can have important implications for managing larger and irregular networks such as the lower Bear River water system.

**Bear River Network.** Here, RANK is applied to identify promising locations for agricultural to urban water transfers and water management actions in the lower Bear River water system of Idaho and Utah. The Bear River watershed comprises 7,500 square miles of agricultural, urban, federal, and state lands in southeastern Idaho, northeastern Utah, and southwestern Wyoming and is the largest tributary of the Great Salt Lake with an average annual inflow of 1.2 million acre feet (Mesner and Horsburgh 2012). The primary water uses in the basin are for agriculture, urban, industrial, power generation, recreation, and the environment.

The lower Bear River system used in this analysis stretches from Southeastern Idaho to the Great Salt Lake, Utah. The Utah Division of Water Resources developed a Bear River simulation model to examine water system sustainability over a 50-year historical record and plan for future infrastructure (new dams and diversions) as well as operational changes (reservoir reoperations, water conservation, and reallocations among demand sites; Utah Division of Water Resources, 2004). The model represents the system with 55 nodes (10 reservoirs [6 existing and 4 proposed], 11 urban, agricultural, and environmental service areas, 13 flow junctions) and 73 linkages (Figure 3). The system’s largest environmental wetland service area is the Bear River Migratory Bird Refuge while other incidental riparian water uses occur at various junctions along the main stem of the Bear River.

As Utah’s urban population continues to grow along the Wasatch front in Salt Lake, Davis, and Weber counties, planners project the need to transport Bear River water to these areas (Mesner and Horsburgh 2012). Water transfers from agricultural to urban uses could change how the water system functions. The water system is complex and managers have numerous possible management options and ways to implement changes. Some questions managers need to address are from what agricultural areas should managers transfer water to meet future urban demands and where might it be appropriate to build additional dams, remove existing dams, implement conservation measures, develop new local resources, monitor flows, or protect environmental and ecosystem services? Applying RANK can help screen promising options on which to conduct further detailed study with simulation models.

Running RANK for the Bear River network gives a parallel coordinate plot where the (i) extracted node axes are ordered and grouped from left to right by reservoirs, service areas, junctions, sources and sinks, and (ii) traces showing extracted centrality values for these node groups are similarly colored (respectively, blue, orange, purple, red, and green). Visual inspection of the plot identifies reservoirs and junctions as the extracted node axes where traces have the largest and most number of drops in extracted centrality values and indicates these nodes are the most topologically significant. Reservoir removal typical causes the centrality of a single green trace to drop to zero which is the reservoir’s corresponding evaporation sink. In contrast, traces are generally flat (few drops) across the abscissa axes that represent extracting service area, source, and sink nodes and show these nodes have little topological significance. Visual and qualitative observations allow the user to generally classify nodes but do not indicate which specific nodes are most or least stable, topologically significant, and redundant. Another drawback to visual inspection is that the order of the axes greatly impacts interpretation of the plot.

RANK’s calculating of pairwise differences among centrality values and ranking of the nodes provides users with quantitative performance metrics of stability, topological significance, and redundancy for each node. Additionally, the ranking calculations control for the axes order. The most stable nodes are sources like the headwaters of the Bear, Blacksmith Fork, and Little Bear rivers while the most unstable nodes are the most downstream nodes that are connected to only one other node (Figure 5 and Table 2). The two most topologically significant nodes—the junctions J32-88 (potential Bear River diversions to Wasatch Front) and J16-17 (existing Bear River diversions to Bear River Canal Company and Box Elder County agriculture) —serve as upstream conduits that connect portions of the network to the main stem of the network (Figure 5 and Table 2).

The topologically significant junctions J32-88, J16-17, and J5-7 (Bear River diversions to Cache Valley) each have a total of 5 immediate links to/from other nodes whereas the less topologically significant Corrine junction has seven immediate connections. The Corrine junction does not connect as large a group of nodes to the main network nor is it as far upstream as the three most topologically significant junctions. These observations show that several factors beyond link density affect node topological significance.

The lower Bear River has several highly redundant node pairs (Table 3). Nodes with 100% redundancy have all of the same connections. Wasatch Service Area and Evaporation from Davis Reservoir each have the same single connection to Davis Reservoir. Cache Agriculture and Cache Urban are 98% redundant as these nodes link to the exact same nodes except that Cache Urban can additionally receive water from Q6-Groundwater whereas Cache Agriculture does not. South Cache Agriculture and South Cache Valley Urban are 95% redundant because they both connect to the same diversion point on the Little Bear River (J48-49) and return flow to the same downstream node (J45-51) but the South Cache Agriculture can additionally divert water from the Blacksmith Fork river below Millcreek Reservoir.

**Discussion**

RANK can identify and rank the most stable, topologically significant, and redundant nodes in a network as well as guide water managers in the further study of their water systems to inform water management decisions. Unstable nodes are more likely to be nodes with fewer connections and located towards the downstream portion of the network. Topologically significant nodes are most likely junctions that serve as flow-through points that connect branches of the network to the rest of the network. Redundant node pairs are nodes that connect to the same other nodes and are usually geographically near each other. Nodes with more connections are less likely to be redundant because there is a lower likelihood that another node will have the same set of connections.

These characteristics are evident in the analysis of the Bear River network. The most unstable nodes are Evaporation from the Bird Refuge, Spill to the Great Salt Lake, and the Great Salt Lake, which are only connected to one node and are located at extreme downstream points. The most topologically significant nodes are Junctions 32-88 (proposed diversions to the Wasatch Front), Collinston, and 16-17 (existing Bear River diversions to Bear River Canal Company and Box Elder County agriculture); these junctions are all centrally located and serve to connect different branches of the network. Several of the more topologically significant locations, such as Junctions 5-7 (diversion to Cache Valley), 48-49 (diversions to South Cache Valley), and Corinne and Hyrum, Willard, and Cutler Reservoirs (all ranking within the top ten), are on primary tributaries before the tributary confluences with the main stem.

For the redundancy analysis, the focus is on node pairs that are of the same type (reservoirs, service areas, junctions, etc.) as these pairs share similar management options. The service areas of Cache Valley Irrigation and Cache Valley New M&I are redundant as are South Cache Irrigation and South Cache M&I; managers can use this redundancy to reroute water in the event of system failure at one node. Redundancy, therefore, identifies operational flexibility in the system. At the same time, the redundancy results also identify system components that serve identical (or nearly identical) functions; managers can use these results to identify nodes that if removed from the system, will save money and other resources but otherwise have little effect on the overall system performance.

Removing nodes that are redundant or not topologically significant will have little effect on network connectivity. Retaining or creating a redundant pair of nodes means that in the event of one node’s disconnection (e.g., pipe break, pump malfunction, etc.), managers can still route water through the other node. Conversely, losing a topologically significant node will affect many other nodes and may mean water cannot be routed to the desired destination.

Managers can use individual or combinations of the RANK performance metrics to improve a simulation model’s representation of the physical system or identify priority management scenarios to study with more detailed model runs. For example, collect additional data and improve network resolution at topologically significant nodes, disaggregate nodes with low stability and high topological significance, or group redundant nodes and nodes with low topological significance to reduce data collection efforts and make the model simpler. Potential management scenarios to study with more detailed simulation modeling can include source nodes for agricultural to urban water transfers as well as locations to build or remove dams or promote water conservation measures. Additionally, locations to develop new alternative supplies, monitor flows, or protect environmental resources. Here, RANK results are further explained relating to develop scenarios of management interest to test with more detailed simulation modeling.

**Potential Water Transfers.** Managers can identify potential sources of water transfers as nodes with high redundancy and low topological significance. For example, RANK results for the lower Bear River system show the Cache Valley Irrigation service area is redundant with the Cache Valley Urban service area. This redundancy suggests that Cache Valley Irrigation has little influence on network connectivity. Its topological significance rank is 42 (out of 55 total nodes) and further emphasizes the redundancy result. Similarly, the South Cache Irrigation service area is highly redundant with the South Cache Urban area and the South Cache Irrigation has a topological significance rank of 43. These results suggest transferring water from South Cache Irrigation to South Cache Urban will have little effect on network connectivity. Together, the redundancy and topological significance metrics suggest the Cache Valley and South Cache service areas may be good candidates for sources of agricultural to urban transfers if the goal of the transfer is to leave intact the connectivity of the remaining parts of the lower Bear River water system. The RANK results for the lower Bear River system identify the two service areas as promising sources of water transfers to further test and explore with more detailed simulation modeling.

**Water Management Scenarios.** In addition to identifying promising sources of water transfers, RANK results can inform many other water management decisions for further testing in a simulation model. Existing or proposed reservoirs with high topological significance should be retained or further studied as potential projects. Reservoirs with low topological significance are candidate sites for dam removals. For example, in the Bear River network, Hyrum Reservoir has a topological significance rank of 7/55 which is the highest of all the reservoirs and suggests that removing this dam would have a substantial effect on network. Willard Bay and Idaho reservoirs have high topological significance values indicating their removals would significantly impact service areas and other nodes in the Bear River network. Additionally, the proposed Washakie, Millcreek, and Mainstem reservoirs have mid-ranking topological significance values (17, 23, and 27, respectively) which suggest that building these reservoirs would impact how other nodes are connected and interact with each other.

Water supply to unstable nodes is easily affected by changes at other network nodes. At unstable nodes, managers should implement water conservation measures, develop new alternative supplies, and monitor flows. In the Bear River system, the least stable service areas are the Weber Basin, Wasatch Front, and the Bird Refuge. These service areas would benefit most from conservation and alternate supplies particularly in times when conditions occur such as low surface water availability, limited reservoir storage, or breaks in water transmission lines. In support of this observation, managers of the Weber Basin Water Conservancy District, who oversee the Weber Basin Project, have steadily promoted water conservation programs over the last decade (Weber Basin Water Conservancy District, 2010). Additionally, managers should also monitor water flows at unstable nodes like the Spill to Great Salt Lake (54/55) and Great Salt Lake (53/55) because these nodes can receive variable flows subject to activities at many other nodes in the network.

Managers should protect nodes with high instability and high topological significance and that provide environmental services because degradation or removal of these areas will negatively impact ecosystem services at the location and other nodes in the network. RANK results can also identify which environmental areas would benefit most from further management. The Bird Refuge, a critical environmental site both in the Bear River system as well as the entire region, has a mid-ranking topological significance value (16/55) but ranks 48/55 for instability in the network which requires it to have very active management. The Bear River network’s fourth-most topologically significant node, J5-7 Diversions to Cache Valley (3/55), is also a critical riparian service area; it is important to protect this site both to preserve its ecological services as well as connected services throughout the water system.

For the lower Bear River, RANK rankings of node stability, topological significance, and redundancy can identify potential reservoir locations, service areas most in need of conservation and alternate supply development, as well as environmental areas to protect. Managers can use the RANK results to develop water management scenarios to test with more detailed simulation modeling.

**Future Work.** With RANK, the stability, topological significance, and redundancy of network nodes are determined based on the unidirectional links from upstream to downstream nodes. The analysis assumes that each node is equally important and there is no limit on the magnitude of flows through links. For example, the analysis assumes reservoir evaporation has the potential to have the same influence on the network as the reservoir inflow or release.

Further analysis should consider variable attributes of each link such as different flow magnitudes or flow capacities (volume or flow rate). When considering variable link attributes, nodes may no longer be redundant if capacity constraints or other considerations limit flow through a link. These variable link attributes could be considered in RANK by including in the calculation of extracted centrality a factor like minimum or maximum flow capacity, average flow, or another measure that describes the pathway of links that connect each pair of nodes. For example, use a distance of three and minimum flow capacity of 5 cms to calculate centrality for two nodes that are three links apart and have intermediary links with capacities of 10, 5, and 7 cms. Thus, centrality becomes a function of both the distance between nodes and the attribute values associated with the path of links connecting the node pair. However, even without flow magnitude, the RANK analysis provides information regarding the stability, topological significance, and redundancy of nodes in a water resources network and can help managers significantly narrow the scenarios they consider with detailed simulation modeling.

**Conclusions**

Modeling complex water resources networks requires significant computational effort. These difficulties are exacerbated when removing individual nodes to study effects on other nodes in the system. In this paper, a Ranking Automation for NetworKs (RANK) tool was developed to identify and rank key network nodes that are (1) stable: their roles do not depend on the existence of particular nodes, (2) topologically significant: when removed or added to the network, these nodes cause other nodes to be unstable, and (3) redundant: node pairs that have similar connections. Node ranks are calculated by taking pairwise differences in centrality values on the parallel coordinate plot that represents network-wide effects of node extractions. These paired differences further control for the ordering of axes on the plot as well as quantify the visual and qualitative interpretation of the plot.

RANK was applied to two small illustrative networks as well as the larger 55-node, 73-link lower Bear River water system. From the analyses, it was determined that unstable nodes generally have very few connections and are located downstream while topologically significant nodes are in the upstream portion of the network and connect large branches of the network. Redundant nodes typically have few but identical connections and similar geographic locations.

The ranking of nodes for stability, topological significance, and redundancy within a water resources network can further help inform water system planning and management. Nodes with high redundancy and low topological significance are candidate sources for water transfers to urban areas. Further, consider (a) removing dams at reservoir sites with low topological significance, (b) building or expanding dams at sites with high topological significance, (c) adopting conservation measures or developing new alternative supplies at unstable service areas, (d) monitoring flows at unstable and topologically significant nodes, and (e) protecting environmental features at unstable and topologically significant locations. RANK identifies the key nodes within a water resources network and allows analysts to narrow the set of manager scenarios to study with further time-intensive modeling and sensitivity analysis.

**Acknowledgements**

This research was funded by Utah Mineral Lease funds. The authors thank the three anonymous reviewers for constructive feedback on the first draft.

**References**

Albazzaz, H., Wang, X.Y., Marhoon, F. (2005). “Multidimensional visualization for process historical data analysis: a comparative study with multivariate statistical process control.” *J. of Process Control,* 15(3), 285-294.

Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet 6 for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Choi, H., Heejo, H., Hyogon, K. (2009). “Fast detection and visualization of network attacks on parallel coordinates.” *Computers & Security,* 28(1): 276-288.

Edsall. R.M. (2003). “The parallel coordinates plot in action: design and use for geographic visualization.” *Computational Statistics & Data Analysis*, 43(4): 605-619.

Harou, J., Pinte, D., Tilmant, A., Rosenberg, D. E., Rheinheimer, D., Hansen, K., Reed, P., Reynaud, Medellin-Azuara, J., Pulido-Velazquez, M., Matrosov, Padula, and Zhu, T. (2010). "An open-source model platform for water management that links models to a generic user-interface and data-manager." 2010 International Congress on Environmental Modelling and Software 'Modelling for Environment’s Sake', Fifth Biennial Meeting, Ottawa, Canada.

Huh, Myung-Hoe and Park, D.Y. (2008). “Enhancing parallel coordinate plots.” *Journal of the Korean Statistical Society*, 37(2008): 129-133.

Inselberg, A. (1985). “The plane with parallel coordinates.” *The Visual Computer* 1, 69-91.

Loucks, D.P., van Beek, E., Stedinger, J.R., Dijkman, J.P.M., Villars, M.T. (2005). Water resource systems modeling: Its role in planning and management. *Water Resources Planning and Management: An Introduction to Methods, Models and Applications.* UNESCO, Paris, France.

Mesner, N. and Horsburgh, J. (2012). “Watershed Description.” *Bear River Watershed Information System,* < <http://www.bearriverinfo.org>> (Nov. 25, 2012).

Rosenberg, D.E. (2012). Visualizing high dimensional data. CEE 6930 Hydroinfomatics course materials. <https://usu.instructure.com/courses/127332/wiki/lecture-materials> (Dec. 12, 2012).

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 2003 Nov 13(11):2498-2504.

Singer, Y. and Greenshpan, O. (2009). Network visualization and analysis with parallel coordinates. *Parallel coordinates: visual multidimensional geometry and its applications*, A. Inselberg, ed., Springer Dordrecht, New York, N.Y., 460-475.

Utah Division of Water Resources. (2010). Map of Bear River simulation area.

Utah Division of Water Rights. (2004). Bear River Basin: planning for the future. <<http://www.water.utah.gov/planning/SWP/bear/bearRiver-1A.pdf>> (Dec. 14, 2012).

Weber Basin Water Conservancy District. (2010). Water Conservation Plan. <http://www.weberbasin.com/conservation/images/docs/conservationplan2010.pdf>> (Apr. 24, 2014).

Wegman, E.J. (1990). “Hyperdimensional data analysis using parallel coordinates.” *J. of the American Statistical Association,* 85(411): 664-675.

Table 1. Terms and definitions used in RANK.

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Adjacency Matrix | A two-dimensional square matrix that defines the links between nodes in a network. Columns and rows enumerate the network nodes. An element in this matrix with a value of one indicates there is a link from the node in the corresponding row to the node indicated by the column. |
| Connectivity Matrix | Square matrix identical in structure to the adjacency matrix where elements specify the minimum number of links from the node indicated by the row to the node indicated by column. Adjacent nodes have a connectivity value of 1. Unconnected nodes have a connectivity value of infinity. |
| Stability | Measure of how much the extracted centrality value for a node changes across extracted networks. |
| Topological Significance | Measure of how extracting a node affects the stability of other nodes in the network. |
| Redundancy | Measure of connection similarity between a pair of nodes. |

Table 2. Most and least stable and topologically significant nodes in the Bear River network.

|  |  |  |
| --- | --- | --- |
| **Rank** | **Stability** | **Topological Significance** |
| 1 | All Equal with Rank 1: Reach Gain, Groundwater Import, Malad River, Surplus from Weber, Q15 Culter Gain, Q46 Little Bear RiverQ46, Q40 Blacksmith Fork, Q6 Groundwater, and Q1 Flow from Bear Lake | Junction 32-88 (potential Bear River diversions to Wasatch Front) |
| 2 | Collinston |
| 3 | Junction 16-17 (Bear River diversions to Bear River Canal Company and Box Elder County agriculture) |
| 53 | Great Salt Lake | Spill to Great Salt Lake |
| 54 | Spill to Great Salt Lake | Evaporation from Mainstem |
| 55 | Evaporation from Bird Refuge | Great Salt Lake |

Table 3. Highly redundant node pairs in the lower Bear River network.

|  |  |  |
| --- | --- | --- |
| **Redundancy Value** | **Node 1** | **Node 2** |
| 100% | Wasatch Service Area | Evaporation from Davis Reservoir |
| 98% | Cache Agriculture Service Area | Cache Urban Service Area |
| 95% | South Cache Agriculture Service Area | South Cache Urban Service Area |