Sparse Matrix Operations in Social Networks

Kritika Prakash , Sai Sri Datta , Abhijeet Mishra, Sourav Chanduka

Introduction to Sparse Matrices

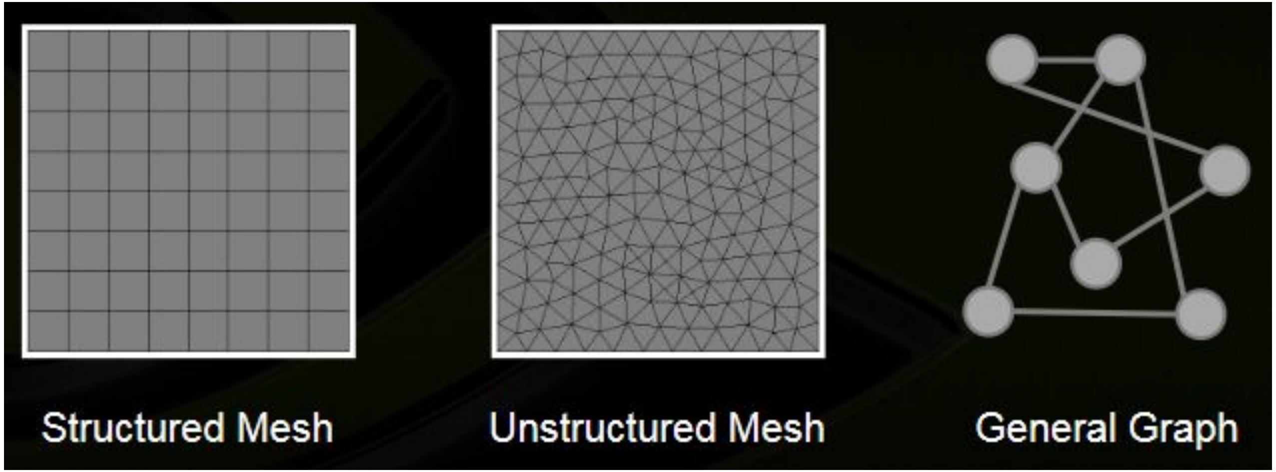
Sparse matrices are at the heart of Linear Algebraic Systems. It is difficult to use very large high-dimensional sparse matrices in memory and do operations on them. Using the right representation can significantly reduce the amount of memory required for data storage and time required for the operation.

Sparse Matrices have finite number of Non-Zero entries.(Which denote Connectivity like Finite Element Meshes , Social Networks ..)

Three basic types of Sparse Matrices:

a) Structured Mesh
b) Unstructured Mesh
c) General Graph

Introduction to Sparse Matrices



Social Network Analysis

1. Components of a social network:
   a. Actors (indivisible) (nodes)
   b. Relationships between the actors (ties)
2. Types of social networks:
   a. **Ego-centric networks:** connected with a single node or individual. For example, you, the node, connected to all your close friends.
   b. **Socio-centric networks :** closed networks by default. Example - children in a classroom or workers inside an organization.
   c. **Open-system networks:** the boundary lines are not clearly defined.

Social Network Analysis

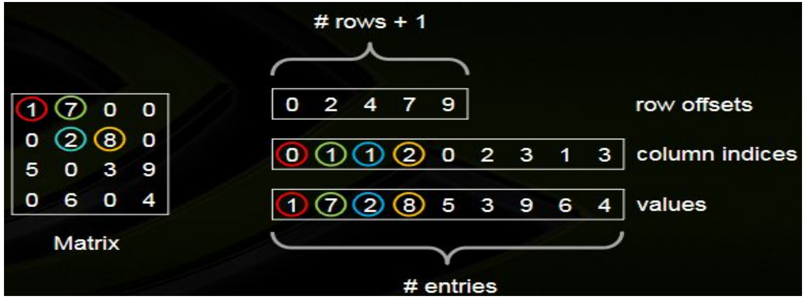
1. Relationships or ties could be directional or non-directional.
2. These ties could also be classified as strong and weak.
3. Actors participate in relationships with varying degrees of closeness, or connectedness.
4. Another important factor is the distance between two nodes.
5. Social Network theory is closely linked to Complexity theory, systems thinking and Graph theory.
6. It overcomes the limitations of statistics as it takes into account the relationships between entities.

Sparse Matrices in Social Networks

1. It is possible and convenient to represent information about social networks in the form of matrices.
2. Representing the information in this way also allows the application of mathematical and computer tools to summarize and find patterns.
3. The matrices used in social network analysis are frequently "square."

Storage Formats

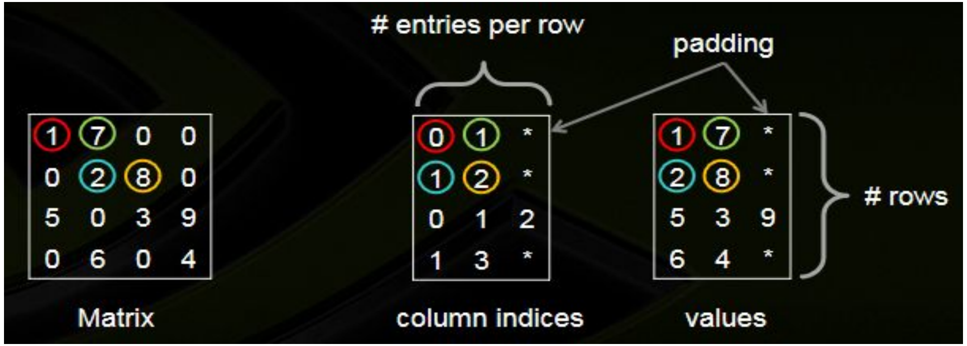
Sparse Matrix Storage Formats :

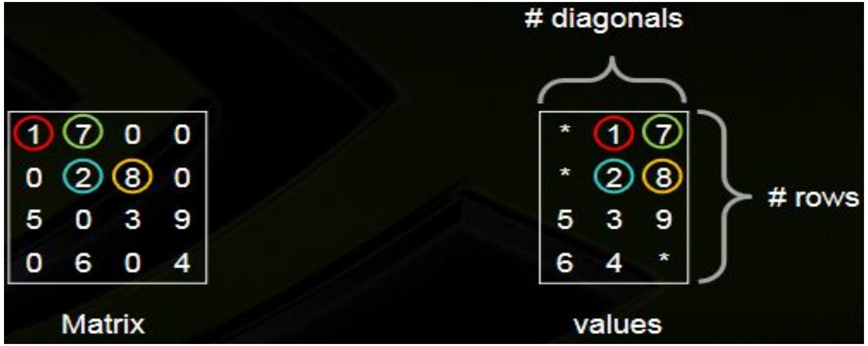


1.) Coordinate

2.) Compressed sparse row

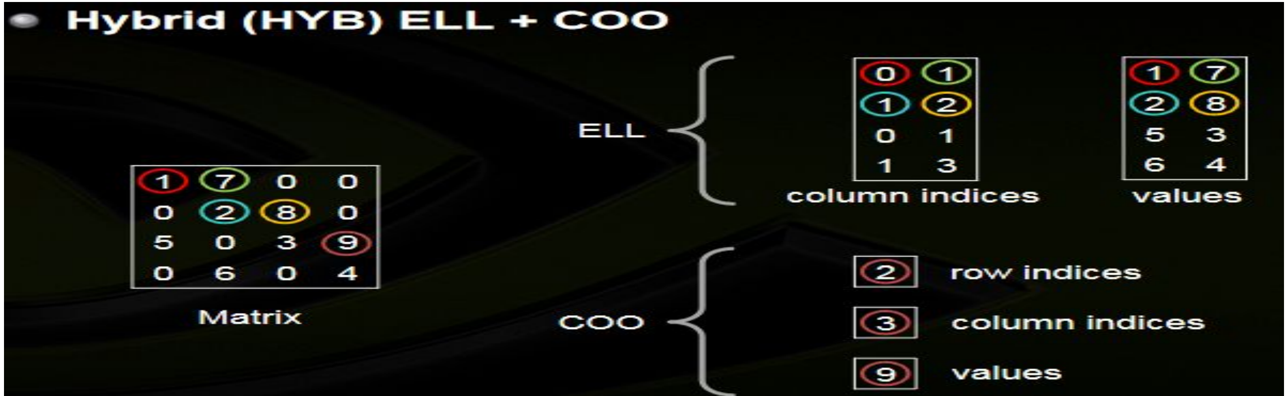
Storage Formats





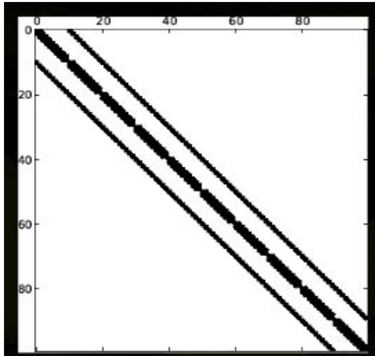
1.)ELLPACK

2.)Diagonal



Storage Format Comparison

Structured Mesh

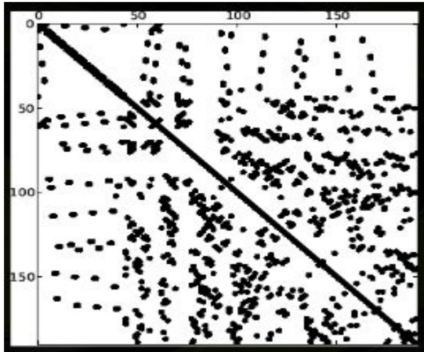


Matrix

Format	float	double
COO	12.00	16.00
CSR	8.45	12.45
DIA	4.05	8.10
ELL	8.11	12.16
HYB	8.11	12.16

Bytes per Nonzero Entry

Unstructured Mesh

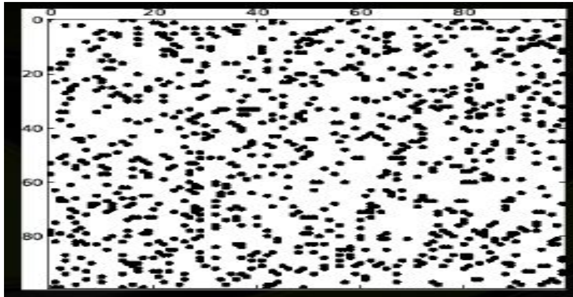


Matrix

Format	float	double
COO	12.00	16.00
CSR	8.62	12.62
DIA	164.11	328.22
ELL	11.06	16.60
HYB	9.00	13.44

Bytes per Nonzero Entry

Random Mesh



Format	Float	Double
COO	12.00	16.00
CSR	8.42	12.42
DIA	76.83	153.65
ELL	14.20	21.29
HYB	9.60	14.20

The Adjacency Matrix

The most common form of matrix in social network analysis is a very simple square matrix with as many rows and columns as there are actors in the network. The "elements" or scores in the cells of the matrix record information about the ties between each pair of actors. All ties are binary (0 or 1, where 1 means friends and 0 means not friends).
Symmetric adjacency matrix: Social distance is symmetric
Non-symmetric adjacency matrix: Social distance is non-euclidean

This is the most basic Representation of data in social Networks. Here, the values in the Main diagonal are meaningless.

	Bob	Carol	Ted	Alice
Bob	---	1	1	0
Carol	0	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---

Sparse Matrix Operations

Computational Complexity

The computational complexity of sparse operations is proportional to the number of **nonzero** elements in the matrix. It also depends linearly on the **row size m** and **column size n** of the matrix, but is independent of the product m*n, the total number of zero and nonzero elements.

1. Permute
2. Block (row and column partition)
3. Image
4. Transpose
5. Addition
6. Multiplication
7. Eigenvectors
8. Inverse
9. Aggregation (collapse)
10. Iterative Solvers

1. Permuting a Matrix

It is helpful to rearrange the rows and columns of a matrix so that we can see patterns more clearly. Shifting rows and columns is called "permutation" of the matrix.

Use case: Rearrange the rows and columns of the sparse matrix to **enhance classification**

Challenge: A good permutation along one dimension might not be good along the other dimensions.

Example: Both males and both females are adjacent in the matrix.

	Bob	Carol	Ted	Alice
Bob	---	1	1	0
Carol	0	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---

Adjacency Matrix

	Bob	Ted	Carol	Alice
Bob	---	1	1	0
Ted	1	---	1	1
Carol	0	1	---	0
Alice	0	1	0	---

Permuted Matrix

2. Blocks of a Matrix

Each colored section is referred to as a *block*. Blocks Are formed by partitioning the matrix along each Dimension.

Partition ~ super-node ~ social role ~ block

Use case: To understand how some sets of actors are "embedded" in social roles, contexts or in larger entities.

Challenge: Examining only the positions or roles. Finding Meaningful roles by dividing into blocks.

Block density matrix →

(Proportion of all ties present within a block)

	Bob	Ted	Carol	Alice
Bob	---	1	1	0
Ted	1	---	1	1
Carol	0	1	---	0
Alice	0	1	0	---

	Male	Female
Male	1.00	0.75
Female	0.50	0.00

3. Image of a Block of a Matrix

Block image or *image matrix* is used to summarize the information further. (threshold density =0.5625)

Image = 1 , if block density >= threshold density

Image = 0 , if block density < threshold density

using overall mean density as the threshold

	Male	Female
Male	1.00	0.75
Female	0.50	0.00

Use case: Images are powerful tools for simplifying the presentation of complex patterns of data.

Challenge: Finding meaningful images of the sparse data

	Male	Female
Male	1	1
Female	0	0

4. Transposing a Matrix

The degree of similarity between an adjacency matrix and its transpose is one way of summarizing the degree of **reciprocity of ties**.

Use Case: Reciprocity of ties can be a very important property of a social structure because it relates to both the balance and to the degree and form of hierarchy in a network.
Challenge: Constraints like symmetric matrix doesn’t always show actual reciprocity of ties.

Example: “Follow” feature in Instagram (Unidirectional, Asymmetric), “Friend” feature in Facebook (Bidirectional, Symmetric)

5. Addition of Sparse Matrices

Matrix addition is most often used in network analysis when we are trying to simplify or reduce the complexity of multiplex (multiple relations recorded as separate matrices or slices) data to simpler forms.

Use Case: If we had a symmetric matrix that represented the tie "exchanges money" and another that represented the relation "exchanges goods" I could add the two matrices to indicate the intensity of the exchange relationship.

Challenge: Adding meaningful relations helps in better analysis.

6. Multiplication of Sparse Matrices

When we multiply an adjacency matrix to itself, i.e., $A*A=A^2$, it counts the number of pathways between two nodes that are of length two. Similarly A^n matrix will count the number of pathways between two nodes of length *n*.

Use case: Individual actor's positions in networks are also usefully described by the numbers and lengths of pathways that they have to other actors. Actors who have short pathways to more other actors may me more influential or central figures.

Challenge: Parallelism, to make the operation faster

7. Eigenvectors of a Sparse Matrix

Use case: Finding community structure in networks using the eigenvectors of matrices.
A community is a groups of vertices with a higher-than-average density of edges connecting them.
Challenge: Parallelism, to make the operation faster and space-optimized
Arnoldi iteration Algorithm: (typical large sparse matrix algorithm)
It does not access the elements of the matrix directly, but rather makes the matrix map vectors and makes its conclusions from their images.
(approximation function)

8. Finding the Inverse of a Matrix

Inverse : Inverse exists only when determinant of a given matrix is non-zero. And the matrix is a square matrix. Inverse Matrix is used in equations like A*X= B. Matrix is singular if it's determinant is 0. Inverse can be calculated using Elementary row operations or Minors, Cofactors and adjugate.

Pseudo-inverse : Denoted by $\mathbf{A^+}$. If the columns of a matrix are Linearly independent pseudo matrix is given by $\mathbf{A^+ = (A^T \cdot A)^{-1} \cdot A^T}$. If the rows are Linearly independent pseudo matrix is given by $A^+ = A^T \cdot (A \cdot A^T)^{-1}$. If rows and columns are linearly independent pseudo matrix = inverse of matrix.

8. Finding the Inverse of a Matrix

Singular value decomposition : It is used for dimension reduction. Given input matrix can be represented as product of 3 matrices. **A** = **UΣV*** Dimension of A is m*n , dimension of U is m*r dimension of sigma is r*r dimension of V is r*n.

U has left singular vectors and V has right singular vectors . sigma is a diagonal matrix whose diagonal elements are called singular elements and are arranged in descending order. For a given Real matrix there is unique possibility of writing it as product of 3 matrices.

Columns of U and V are orthogonal. And sum of squares of elements in a column is 1.

So a data Matrix is taken and SVD is applied to it , from which we obtain how much a user corresponds to a particular type in the input matrix from matrices U , V and sigma.

9. Aggregating a Matrix (Collapse)

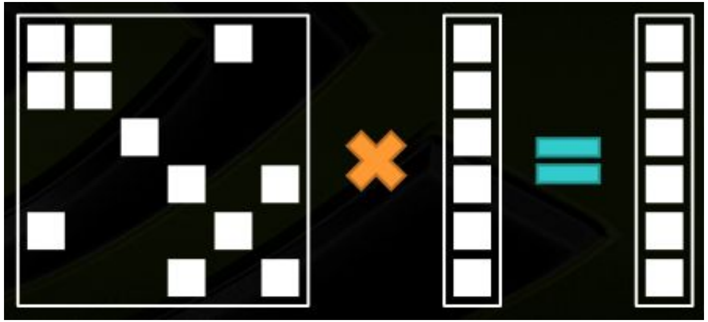
Collapse allows you to combine rows and/or columns by specifying which elements are to be combined, and how.

We might select, for example, to combine columns, and rows by taking the average of the values (we could also select the maximum, minimum, or sum).

It creates a new matrix that has collapsed the desired rows or columns using the summary operation selected.

Use case: Creating meaningful information through the combination of existing data fields

10. Iterative Solvers of a Sparse Matrix



Sparse Solvers :

- Solve sparse linear system (A*X=b).
- There are methods like Direct and Iterative.

<u>Iterative Solvers converge to exact Solution :</u> Iterative solvers are Scalable, Use low Memory and amenable to parallelism compared to Direct solvers.

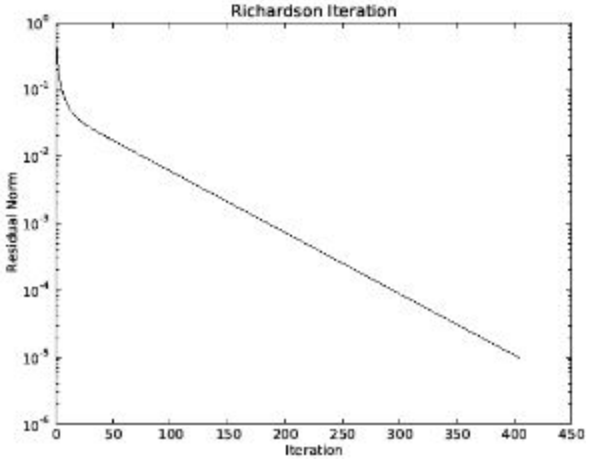
They measure error through residual :

- Residual = b-A*X
- Residual = A*(X'-X) = A*error
- We stop the process when ||b-A*X|| < Tolerance

10. Iterative Solvers of a Sparse Matrix

Several Iterative methods are :

- Jacobi
- Gauss-Seidel
- Conjugate Gradient (CG)
- Generalised Minimum Residual (GMRES)



Richardson Iteration (Example of Iteration):
r = b - A*X

While norm(r) > Tolerance (Norm is square root over sum of square of elements) :
X = X+omega*r
r = b - A*X
Error decreases with iteration

Applications in Social Networks - Examples

- The concept of cliques (representable as a matrix) finds its way in Social Networks as well as Clustering.
- Finding community structure in networks using the eigenvectors of matrices.
- Used to compute clusters or “communities” in graphs representing social networks.
- Used to predict user behaviour depending on the “similarity” of behaviour among users.

The Small World Phenomenon

- The idea that we’re all connected by just “six degrees”—six other people
- The goal of the experiment was to find short chains of acquaintances linking pairs of people in the United States who did not know one another.
- Exhibited by artificial as well as natural networks
- **Network Models**: Watts and Strogatz showed that, beginning with a regular lattice, the addition of a small number of random links reduces the diameter—the longest direct path between any two vertices in the network—from being very long to being very short.
- **Overview:** Build a representation of a network using sparse matrices and then using the sparse matrix operations, the diameter of the network is computed.

Sparse Matrix Representation for Web Opinions

- Based on the work by Dr.L.Venkateswara Reddy & B.R.Dileep Kumar
- Due to the advancement of Web 2.0 technologies, a large volume of Web opinions is available on social media sites such as Web forums and Weblogs.
- A sparse matrix representation was used for web opinions. An algorithms for matrix generation was proposed using a vector of forum threads.
- **Overview:**
  - Preprocessing of text in forums - tokenize and generate term frequencies.
  - Vector Gathering - convert the terms and their frequencies to vector form.
  - Matrix generation - A sparse matrix was generated from the vectors which was used to represent web opinions.

Using Sparse Matrix Kernels for Data Mining

- Many data mining algorithms rely on eigenvalue computations or iterative linear solvers.
- **Text Retrieval** :In Latent Semantic Indexing (LSI), a term-by-document matrix is projected to a smaller dimensional space using by computing the truncated singular value decomposition SVD of the matrix, and retrieval is performed by projecting the query onto the same space.
- **Eigenface Approximation (EA):** The algorithm systematically coarsens images to create a multi-resolution hierarchy of the image set and computes co-eigenvectors for the coarsest images; it works its way up, and finally recovers primary eigenvectors for the original images from their approximate co-eigenvectors.

Drawbacks of Sparse Matrix Representation

- Some sparse matrix operations are slower than their dense matrix counterparts due to irregular memory access patterns and to indirection overhead in the sparse data structure.
- Sparse matrix-vector multiplication often runs one to two orders of magnitude below the peak hardware performance.
- Sparse matrix performance strongly depends on the nonzero structure of the matrix, with the worst performance often seen from data mining matrices.
- Sparse matrix generation algorithms need to customized according to the need.

Conclusion

1. Matrices are used in network analysis to represent the adjacency of actors in a network.
2. An adjacency matrix is a square actor-by-actor (i=j) matrix where the presence of pair wise ties are recorded as elements. The main diagonal, or "self-tie" of an adjacency matrix is often ignored in network analysis.
3. Sociograms, or graphs of networks can be represented in matrix form, and mathematical operations can then be performed to summarize the information in the graph. Vector operations, blocking and partitioning, and matrix mathematics, are mathematical operations that are sometimes helpful to let us see certain things about the patterns of ties in social networks.
4. Once a pattern of social relations or ties among a set of actors has been represented in a formal way (graphs or matrices), we can define some important ideas about social structure in quite precise ways using mathematics for the definitions.

Thank you!

Team 11