Interactions toolbox: a brief user’s guide

Jan Šimbera, 3. 10. 2015 (simberaj@natur.cuni.cz)

# Abstract

This document describes how to use the Interactions toolbox for ArcGIS 10. Inputs and outputs are presented, followed by a step-by-step guide to expected use cases. Does not contain detailed documentation about the individual tools; for this, refer to the tool description within ArcGIS.

# License and Compatibility

The toolbox is released under the MIT License. The text of the license is disclosed with the toolbox.

Efforts have been made to keep the toolbox compatible across ArcGIS 10 versions; however, most tests have been performed on ArcGIS 10.2, Windows 7 64-bit. Should any errors occur, please contact the author and enclose as much information about the error and its whereabouts.

# Introduction

The Interactions toolbox was developed to automate working with spatial interaction (flow) data, such as commuting matrices, containing some other useful features such as regionalization. It consists of a number of toolsets:

* Conversion converts between the feature class types used by the toolbox (see the Dataset Types section for details).
* Management modifies the interactions in a systematic way (sorting, summarization, selection...)
* Modeling fits some model to the interactions and computes its characteristics for them.
* Neighbourhood deals with spatial unit neighbourhood concepts and computes neighbour tables.
* Network creates new interaction tables based on a network dataset and a set of places.
* Regionalization performs spatial zone aggregation based on their interactions and computes some of the resulting region characteristics.
* Visualisation creates auxiliary data useful for cartographic outputs of the Regionalization toolset results.

To view some common use cases, skip to part 3. To get detailed information about a single tool, refer to its help inside ArcGIS.

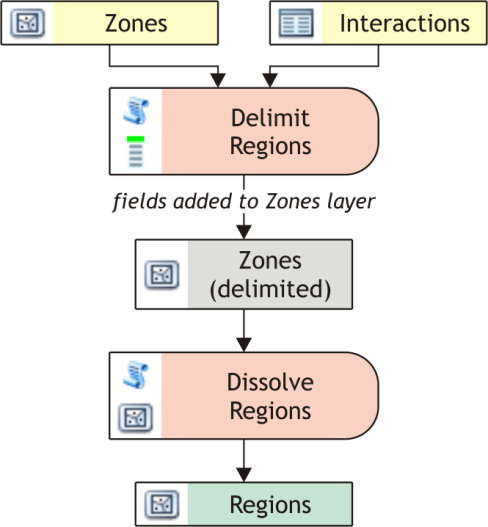
# Dataset Types

This section describes the dataset types used by the toolbox as inputs and outputs. Mostly, they are ArcGIS tables (i. e. datasets without spatial reference) or feature classes (i. e. datasets with spatial contents, such as shapefiles or geodatabase feature classes) satisfying some additional properties (such as having some specified attributes – columns, fields) required for the tools using them to run correctly.

|  |  |
| --- | --- |
|  | **Interaction Table**. Has at least three fields:   * Origin Identifier – a numeric or text field uniquely identifying the interaction origin (source, start); * Target Identifier – a numeric or text field uniquely identifying the interaction target (destination, end); * Strength – a numeric field denoting the interaction strength (weight – such as number of commuters).   Some other fields may also be used by some tools:   * Length – a numeric field denoting the distance along which the interaction occurs.   May be substituted by an Interaction Feature Class |
|  | **Interaction Feature Class**. A line feature class whose shape field does not contain any route information, only less valuable information such as as-the-crow-flies lines connecting origins and targets, and its attribute table satisfies the requirements for Interaction Table.  May be used anytime the Interaction Table is required; when interaction length field is required, this type is usually better as the distance can be computed easily.  May be created from the Interaction Table and Zone Points by the Table to Interactions tool. |
|  | **Connection Feature Class**. A line feature class whose shape field contains route information (exact shape describing the route from the origin to target) and its attribute table satisfies the requirements for Interaction Table.  May be used anytime the Interaction Table is required; when interaction length field is required, this type is usually better as the distance can be computed easily and is more precise; however, in other cases the larger file size may slow down the calculations.  May be created from the Interaction Table, a Network Dataset and Zone Points by the Table to Connections tool. |
|  | **Zone Polygons**. A polygon feature class containing base spatial units (BSUs) that can be used for regionalization (aggregation). Should contain an ID field uniquely identifying the zone (numeric or text) so that the interactions’ sources and targets may be joined to the zones.  Some other fields may also be used by some tools:   * Mass – a numeric field containing the weight of the zone (such as number of inhabitants). * *Name* – a text field containing the name of the zone (such as the administrative name). |
|  | **Zone Points**. A point feature class containing base spatial units’ representative points (such as centroids). The field requirements are the same as for Zone Polygons, from which it may be created using the standard toolbox Feature to Point tool.  Useful for Conversion tools or creating Voronoi polygons, not so much for regionalization as region delimitations can’t be produced from them. |
|  | **Zone Polygons with Region ID Field**. Same as Zones, but also has to contain a Region ID Field that determines the region to which the zone is assigned. Most tools just write the Region ID Field into the input Zones attribute table. |
|  | **Settlement Polygons**. A polygon feature class delimiting built-up settlement areas. The attributes do not matter. |
|  | **Network Dataset**. A built network dataset. |
|  | **Neighbourhood Table**. A table determining feature (such as Zone Polygons) neighbourhood. Is used by most tools to speed up the calculation as its determination is rather costly and the result may be used multiple times.  Must contain these two fields:   * ID\_FROM – the identifier (numeric or textual) of the feature, * ID\_TO – the identifier (numeric or textual) of its neighbour.   The relationship implied by a single row is directed (i. e. having a row A B implies that B is neighbour of A but not necessarily that A is neighbour of B).  May be computed by tools from the Neighbourhood toolset; the basic choice is Administrative Neighbourhood Table. |
|  | **Region Polygons**. A polygon feature class with the region polygons delimited in the regionalization process. Useful for some tools and simpler visualisation, but for some tools the *Zone Polygons with Region ID Field* should be used instead. |

# Common use cases

## Regionalization

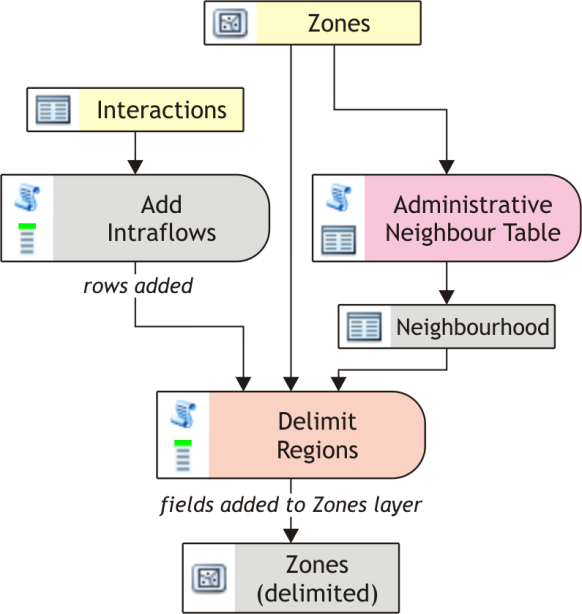
The aim of regionalization is to delimit regions as aggregates of zones (base spatial units, BSUs) that are optimal with respect to some criteria, such as maximum self-containment of interaction flows or minimum mass.

The most basic workflow is displayed to the right. **Required inputs** are:

* *Zone Polygons*
* *Interaction Table*

These should be input into the Delimit Regions tool that does most of the work – determines which zone belongs to which region according to the interactions and, for each zone, writes the region identifier (which usually is the identifier of the most important zone of the region) into the Zone Polygons attribute table under the specified field. Dissolve Regions is used afterwards to gain polygon delimitations of the regions (via a common dissolve procedure) plus some aggregate characteristics of the region.

Besides this basic procedure, some tools may be used to enhance it as shown to the right:

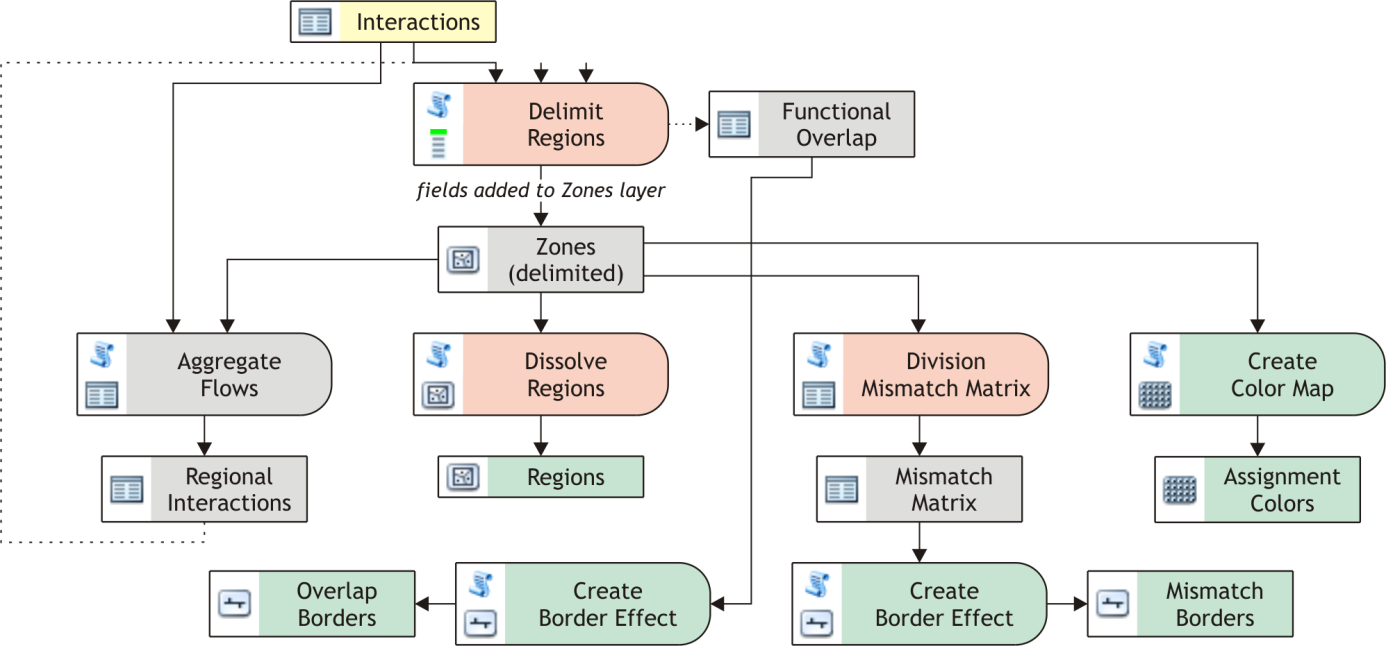
* Administrative Neighbour Table precomputes the Zone Polygons neighbourhood table that can be passed to Delimit Regions as an optional input, which speeds the tool up significantly. As the neighbourhood table can be reused when delimiting regions more than once above a given Zone Polygons layer, it is usually useful to have it precomputed. If the neighbourhood table is not passed to Delimit Regions, the tools automatically calls Administrative Neighbour Table itself.
* For some regionalization algorithms such as TTWA, intrazonal interactions (intraflows representing for example the number of people living and working within the same zone) must be contained within the Interaction Table – otherwise the self-containment levels are not computed correctly. However, some tables do not contain these flows (such as most census commuting tables). Add Intraflows can be used to fix this and add these intraflows to the input interaction table, provided it also gets the corresponding zone layer with a field containing the total number of source/target items (such as the total number of economically active population or jobs in the zone).
* Some tools from the Management toolset described in the section below may also come in handy to modify the input interaction table.

If multi-level (multi-tier) regionalization is desired, it is necessary (besides obtaining the region polygons using Dissolve Regions) to get interactions between regions instead of between their zones. Aggregate Flows can be used for this.

To visualise and present the resulting regionalization, multiple tools are also available as described below and shown on the diagram afterwards:

* Division Mismatch Matrix can be used to analyze differences between two region delimitations, provided they are stored in a single zone layer. The output is a table that assigns to a pair of the region identifiers the total mass of all zones assigned to the first region in the first regionalization and to the second region in the second regionalization (but not vice versa, that forms an another row).
* Create Border Effect can be used to assign the Mismatch Matrix (the result of the tool above) or Functional Overlap (an optional output of Delimit Regions) to the borders of the resulting Region Polygons and create a line feature class, easily visualisable for example by line thickness.
* Calculate Measures can be used to compute some additional optional outputs of the Delimit Regions tool without calculating the delimitation again (otherwise, all the steps are performed the same).
* The most advanced visualisation is the quantitative-qualitative color map. It shows fuzzy region boundaries by mixing base region colors for the zones to show how much they are bound not only to their own region, but to all others.

To create it, multiple steps are necessary:

* + Run Assign Colors on the resulting Region Polygons. This assigns base saturated colors to the regions as a new field with a hexadecimal RGB color code.
  + Join the base color field from the Region Polygons to the Zone Polygons used in the regionalization using Join Field from the standard ArcGIS toolbox (remember to use the zone, not the region, identifier field as the Input Join Field).
  + Run Calculate Measures with the same parameters as Delimit Regions during the regionalization, only turning the Output Assignment Color Field option on and specifying the Zone Color Field as the field created using the Join Field tool. This will create another new field in the Zone Polygons with the resulting color (again as a hexadecimal RGB code).
  + Run Create Color Map on the Zone Polygons. The result is a raster image displaying the assignment colors.

## Gravity Modeling

Another use of the toolbox is to perform gravity modeling and analysis. For this, the Model Gravity Interactions is the main tool; it may also be good to prepare the interactions using some of the tools from the Management toolset (such as Join OD Data to join origin and target masses to the interactions).

To gain interaction length, Table to Connections or Table to Interactions may be used.

## Management Toolset

There are tools in the Management toolset to simplify interaction handling, namely:

* Select Interactions creates a new interaction table (or feature class, retaining geometry) with only interactions satisfying up to three conditions: sufficient absolute strength, sufficiently short length and/or sufficient strength relative to length.
* Select Interactions by Origins creates a new interaction table (or feature class, retaining geometry) with only interactions occurring between origins in the specified zone table or feature class. Useful for subsetting the interactions for exploring a smaller portion of space.
* Add Intraflows and Aggregate Flows have been described in the Regionalization section.
* Join OD Data joins fields from the origins table or feature class to the interactions so that each interactions gets the specified information about both its origin and target.
* Mark Interactions assesses the interaction relevance for their origin and target and marks them by added fields containing their order by strength, strength relative to the strongest interaction of the origin/target, and flow significance by Van Nuffel (binary information).
* Mutualize Interactions joins fields to the interactions that belong to the respective interaction in the opposite direction (so that each flow has now fields with information about its counterflow, which can be summed, compared...)
* Summarize Interactions (not to be confused with Aggregate Flows) marks the zones with sums of their inflows’ and outflows’ strengths (as new fields).